Directed evolution of an industrial *N*-acetyl-amino acid racemase

Scott Baxter

A thesis submitted for the degree of doctor of Philosophy

The University of Edinburgh

2011

Abstract:

The use of stereoselective aminohydrolases (acylases) in kinetic resolutions is a commonly employed industrial route to both L- and D- α-amino acids from *N*-acetylated-DL-starting materials. However, a flaw in this process is the need for repeated racemisation steps of the non-desired enantiomer to achieve yields >50%. A solution to this drawback would be a dynamic kinetic resolution driven by an *in situ* racemisation step that would allow the yield to approach 100% (**Figure 1**).

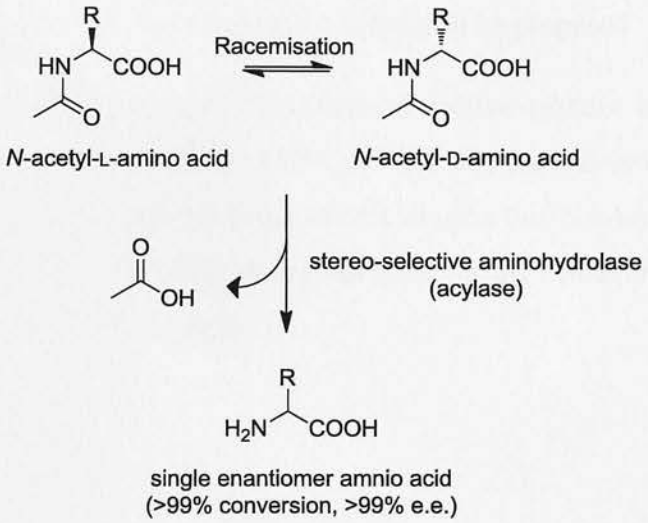


Figure 1. A proposed NAAAR/acylase coupled dynamic kinetic resolution of an enantiopure amino acid.

A cheap and "green" catalyst for this racemisation would be an enzyme such as *N*-acyl amino acid racemase (NAAAR) from the actinobacteria *Amycolatopsis sp. Ts-1-60*. This enzyme requires no organic cofactor, is stable at high temperatures (~60°C) and importantly, shows no racemase activity toward free amino acids. Unfortunately, the activity of NAAAR with *N*-acetyl substrates is low and reported to suffer from inhibition above substrate concentrations of 50 mM, prohibiting its use in NAAAR/acylase coupled dynamic kinetic resolutions under industrial conditions.

In an attempt to remove these limitations, directed evolution has been applied to the *Amycolatopsis Ts-1-60* NAAAR to engineer a variant enzyme with increased activity towards *N*-acetyl substrates. An improved variant NAAAR may allow for a commercial NAAAR/acylase coupled dynamic resolution process. Directed evolution has proven to be a highly versatile and successful tool for protein engineering. However, the ability to screen for improved variants is often technically difficult or time consuming and in the case of NAAAR this is especially true as the

substrate and product are simply enantiomers of an *N*-acetyl amino acid. To overcome this, an enantioselective genetic selection system has been employed to allow the screening of mutagenic NAAAR libraries. After several rounds of selection, a NAAAR variant (NAAAR G291D F323Y) has been isolated with increased racemase activity towards a range of synthetically useful *N*-acetyl substrates. This enzyme has been over-expressed, purified and its characteristics compared to the wild-type and other variants discovered during the evolution process. NAAAR G291D F323Y has been crystallised to 2.71 Å allowing a molecular basis for the increase in catalytic activity to be proposed.

Coupling of this enzyme with a stereoselective acylase has been used to produce enantiopure amino acids in >99% yield, twice the inherent 50% maximum yield of acylase based resolutions. Early results suggest this NAAAR variant has the potential to be employed on a multi kilogram scale for the economical production of enantiopure L- and D- α-amino acids.

Declaration:

I, Scott Baxter, hereby certify that this thesis has been composed by me and that it is a record of my work, and that it has not been accepted in partial or complete fulfilment of any other degree or professional qualification.

Scott Baxter

The University of Edinburgh

2011

Acknowledgements:

Firstly, I would like to thank my PhD supervisors, Dr Dominic Campopiano, Dr Karen Holt-Tiffin, and Dr Ian Fotheringham for their valued input and contribution to this work. I am grateful for all I have learnt and experienced working under their supervision.

I would especially like to thank Dr Ian Taylor for making my industrial placement such an excellent learning experience. I will look back on my time in Cambridge with nothing but fond memories. This is undoubtedly helped by the warm welcome I experienced at Chirotech Technology and the excellent work atmosphere they maintain.

Dr Gideon Grogan must be acknowledged and thanked for his patience, advice and for allowing me to carry out structural biology work under his supervision.

At Ingenza, I would like to thank Ian Archer, Ruben and Fraser for their help and advice with many aspects of my PhD and for allowing me use of their facilities.

My time in Lab 229 has been long, but always enjoyable, especially the last year in which John, Ashley and Ed have brought laughter and a sense of harmony that was previously lacking. I am indebted to Gareth Morrison, Josefin Bartholdson, Venu Bhat, Jonathan Lowther and Marine Raman, who all helped massively during the steep learning curve I experienced. I would like to thank Karin Bodewits for her help, advice and for making me look good ☺. My PhD experience would have been very different without Sam, so I must express thanks for her support and help through it all.

Special thanks are due to Dragomir, Susanna, Neil, Ambur, Tobias, and every other project student who worked in Lab 229. Their endless questions and queries contributed greatly to developing my understanding of protein chemistry/molecular biology and I am truly grateful for this.

Most importantly, I would like to thank my parents - I hope I can repay all you have done for me.

This work has been funded by the BBSRC and Dr Reddys.

Abbreviations:

Enzymes:

α-amino-ε-caprolactam racemase	ACL, EC 5.1.1.15
Alanine racemase	Alr, EC 5.1.1.1
Aspartate aminotransferase	AspTA, EC 2.6.1.1
Aspartate racemase	AspR, EC 5.1.1.13
β-Methylaspartate ammonia lyase	MAL, EC 4.3.1.12
Branched-chain amino acid aminotransferase	BcaT, EC 2.6.1.42
Cystathionine-γ-synthase	MetB, EC 2.5.1.48
D-Amino acid dehydrogenase	D-AAD, EC 1.4.5.1
D-Amino acid oxidase	D-AAO, EC 1.4.3.3
Deoxyribonuclease	DNase, EC 3.1.21.1
dTDP-glucose 4,6-dehydratase	GlucD, EC 4.2.1.46
Formate dehydrogenase	ForDH, EC 1.2.1.43
Glucose dehydrogenase	GHD, EC 1.1.99.10
Glutamate racemase	MurI, EC 5.1.1.3
Glutaryl acylase	GAC, EC 3.5.1.93
Horseradish peroxidase	HRP, EC 1.11.1.7
Thermococcus kodakaraensis polymerase	KOD, EC 2.7.7.7
Lactate dehydrogenase	LDH, EC 1.1.1.27
L-Ala-D/L-Glu epimerase	AEE, EC 4.2.1.-
L-Amino acid dehydrogenase	L-AAD, EC 1.4.1.5
L-Amino acid oxidase	L-AAO, EC 1.4.3.2
Leucine dehydrogenase	LeuDH, EC 1.4.1.9
Lovastatin nonaketide synthase	LNS, EC 2.3.1.161
Mandalate racemase	MR, EC 5.1.2.2
Mannonate dehydratase	ManD, EC 4.2.1.8
Muconate lactonizing enzyme	MLE, EC 5.5.1.1
N-acetyl amino acid racemase	NAAAR, EC 5.1.1.10
Nicotinamide adenine dinucleotide oxidase	NOX, EC 1.6.3.1
Nitrile hydratase	NHase, EC 4.2.1.84
Non-ribosomal peptide synthase	NRPS
N-succinyl amino acid racemase	NSAAR, EC 5.1.1.10
Orthinine transaminase	OTA, EC 2.6.1.13

Penicillin G amidase	PGA, EC 3.5.1.11
Penicillin V amidase	PVA, EC 3.5.1.11
Pyrococcus furiosus polymerase	*Pfu,* EC 2.7.7.7
Phenylalanine racemase	GrsA/GrsB, EC 5.1.1.11
Proline racemase	ProR, EC 5.1.1.4
Pyruvate dehydrogenase	PDH, EC 1.2.4.1
Pyruvate kinase	PK, EC 2.7.1.40
o-succinylbenzoate synthase	OSBS, EC 4.2.1.113
Thermus aquaticus polymerase	*Taq*, EC 2.7.7.7
Triose phosphate isomerase	TIM, EC 5.3.1.1

Chemicals/materials:

(+)-6-aminopenicillanic acid	6-APA
2'-deoxy-*p*-nucleoside-5'-triphosphate	dPTP
2-hydroxy-6-succinyl-2,4-cyclohexadiene carboxylate	2-HSCHC
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	HEPES
5-(dimethylamino)naphthalene-1-sulfonyl chloride	dansyl-Cl
7-amino-3-deacetoxy cephalosporanic acid	7-ADCA
7-aminocephalosporanic acid	7-ACA
8-oxo-2'-deoxyguanosine-5'-triphosphate	8-oxo-dGTP
Acetonitrile	ACN
Acetyl-coenzyme A	Acetyl-CoA
Adenine	A
Adenosine-5'-monophosphate	AMP
Adenosine-5'-triphosphate	ATP
Cell free extract	CFE
Chiral stationary phase	CSP
Coenzyme A	CoA
Cytosine	C
DL-homophenylalanylhydantoin	DL-HPAH
Davis minimal agar	DMA
Davis minimal broth	DMB
Deoxyribonucleic acid	DNA
Deoxyribonucleotides	dNTPS
Dimethyl sulfoxide	DMSO

Double distilled H_2O	ddH_2O
Ethyl methanesulfonate	EMS
Ethylenediaminetetraacetic acid	EDTA
Flavin adenine dinucleotide	$FAD/FADH_2$
Guanine	G
Isopropyl β-D-1-thiogalactopyranoside	IPTG
L-homophenylalanine	L-HPA
Luria Bertani agar	LB agar
Luria Bertani broth	LB broth
Methanol	MeOH
N-acetyl-L-cysteine	NAC
N-carbamoyl-L-homophenylalanine	*N*-carb-L-HPA
Nickel nitrilotriacetic acid	Ni:NTA
Nicotinamide adenine dinucleotide	$NADH/NAD^+$
Nicotinamide adenine dinucleotide phosphate	$NADPH/NADP^+$
o-phthaldialdehyde	OPA
o-succinylbenzoate	OSB
Super optimal culture	SOC
Polyethylene glycol	PEG
Polyethylene glycol monomethyl ether	PEG MME
Pyridoxal-5'-phosphate	PLP
Pyridoxal-5'-amine	PMP
Thymine	T
Triethylammonium acetate	TEAA
Tris(hydroxymethyl)aminomethane hydrochloride	Tris-HCl

Techniques:

Downstream processing	DSP
Dynamic kinetic resolution	DKR
Error prone polymerase chain reaction	epPCR
Fluorescence assisted cell sorting	FACS
High pressure liquid chromatography	HPLC
Immobilised metal affinity chromatography	IMAC
Ion exchange	IEX

Kinetic resolution	KR
Megaprimer PCR of whole plasmid	MEGAWHOP
Polymerase chain reaction	PCR
Reverse phase high pressure liquid chromatography	RP-HPLC
Site directed mutagenesis	SDM
Size exclusion chromatography	SEC
Sodium dodecyl sulphate polyacrylamide gel electrophoresis	SDS-PAGE

General:

Absorbance at 280 nm	Abs_{280}
Amycolatopsis Ts-1-60	*Am. Ts-1-60*
Base pair	bp
Colony forming unit	c.f.u
Column volume	cv
Dalton	Da
Enantiomeric excess	e.e.
Enantioselectivity	E
Escherichia coli	*E. coli*
First order kinetics	FOK
Internal standard	I.S.
Isoelectric point	pI
Melting temperature	T_m
Molar absorption coefficient	ε
Optical density at 600 nm	OD_{600}
Protein databank	PDB
Rate of hydrolysis	$k_{hyrdolysis}$
Rate of racemisation	$k_{racemisation}$
Room temperature	rt
Unit of enzyme activity	U
Wild-type	WT

Chapter 1, Enzymes as biocatalysts for the production of α-amino acids:

1.1, Enzymes as biocatalysts:

Enzymes have been used for several thousand years to carry out simple chemical reactions (biotransformations), for example the fermentation of ethanol from sugar by *Saccharomyces cerevisiae* (baker's yeast)[1]. The science behind these processes was of course a mystery to our ancestors, but today we have the knowhow and ability to use enzymes to carry out a number of chemically exotic and diverse biotransformations. In the previous century an improved understanding of both enzymology and recombinant protein expression has allowed biocatalysis to become an established branch of chemical synthetic techniques[2-7]. It is generally viewed as a "greener" alternative to traditional chemocatalysis due to the mild conditions required for many enzymatic reactions, and this no doubt fuels a large amount of the interest that exists in the field[8, 9]. Chemical feed stocks (both organic and metallic) are dwindling and biocatalysis may help to alleviate some of the demands on these resources[10-12]. However, biocatalysis often struggles to match chemical means in terms of cost, as it is still relatively expensive to develop an effective biocatalyst. For this reason, biocatalysis is mostly used in the realm of pharmaceutical and fine chemicals where the low volume/high value adage exists[13]. Yet, there are exceptions, such as the synthesis of acrylamides from acrylonitriles by Nitrile hydratase (NHase, EC 4.2.1.84) (**Figure 2**)[14].

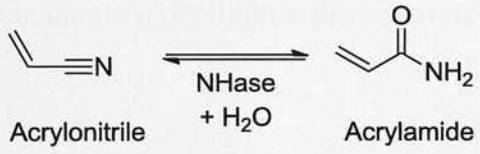


Figure 2. Nitrile hydratase mediated hydrolysis of acrylonitriles to generate acrylamide.

This biocatalytic route to acrylamides proved to be cheaper, more environmentally friendly, and resulted in an improved yield and purity of final product compared to the previous chemo-catalytic route[15]. NHase is only one of a number of commonly used biocatalysts, most of which belong to distinct enzyme classifications dependent on their chemistry. Directed evolution and protein engineering will no doubt help to improve the range of available biocatalysts and possibly overcome the inherent problems (low substrate range, activity and stability) limiting their application[16-18].

This PhD project seeks to develop an economic process to α-amino acids which are useful intermediates in the pharmaceutical and fine chemical industries. To

introduce the subject, examples of reaction types and enzymes that are relevant to the industrial production of these building blocks are given below.

1.2, Types of enantioselective biocatalytic reactions:

In the field of biocatalysis, there are three general reaction types that can be used to synthesise enantio-pure α-amino acids – kinetics resolutions, dynamic kinetic resolution/de-symmetrisations, and asymmetric synthesis reactions.

1.2.1, Kinetic resolutions:

Kinetic resolutions (KRs) are employed to resolve a racemic starting material into a single enantiomer in a 50% yield. The process can be operated such that the desired product is formed during the biotransformation, or that the biocatalyst only transforms (removes) the unwanted enantiomer leaving the desired product un-reacted. In either case, only a 50% yield can be achieved. A typical (and appropriate) example of this process would be the use of an enantioselectivity acylase to de-protect a racemic *N*-acyl amino acid (**Figure 3**). In these reactions, the relative rates of reaction with each enantiomer need to be significantly different to allow for the resolution to proceed with high e.e. If cost allows, it may be possible to racemise or recycle the 'off' enantiomer and repeat the biotransformation to increase the yield above 50%, however the process needs to be repeated 7 times to reach a theoretical yield of 99% making this un-attractive unless the substrate is expensive or in low supply.

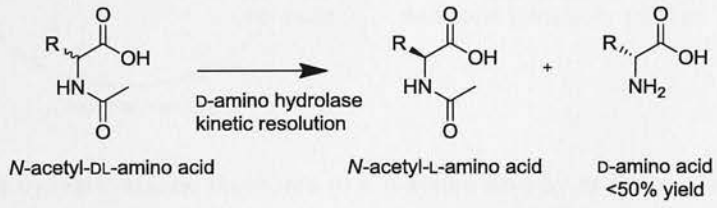


Figure 3. The kinetic resolution of a D-amino acid from racemic *N*-acetyl-amino acid starting material with a D-amino hydrolase.

1.2.2, Dynamic kinetic resolutions/de-symmetrisations:

The second type of biocatalytic reaction is an improved form of the KR which allows for yields greater than 50% to be achieved in a single pot reaction. Dynamic kinetic resolutions (DKRs) employ a racemising agent selective for the substrate (either chemical or enzymatic) to racemise the 'off' enantiomer allowing for it to be continually converted to desired product. Unlike substrate recycling

which can be performed once a KR has reached its 50% yield end point, the racemisation is done continually, *in situ* (**Figure 4A**). The term DKR is applied to resolutions in which the product is formed as the reaction proceeds, allowing the unreacted starting material to be racemised. A form on DKR in which the racemic starting material is converted to a single enantiomer of that material is called a de-symmetrisation. Here, an enzyme is used to transform one enantiomer, and this molecule then chemically converted back to chiral starting materials. As a result, a single enantiomer constantly undergoes racemisation leading to a decrease in its concentration until it has been completely removed (**Figure 4B**). Both processes can lead to a theoretical 100% conversion and 100% e.e. During a DKR, a chemical transformation takes place that involves either bond breaking or bond making, whereas during de-symmetrisation the starting material undergoes no net chemical change other than stereoinversion.

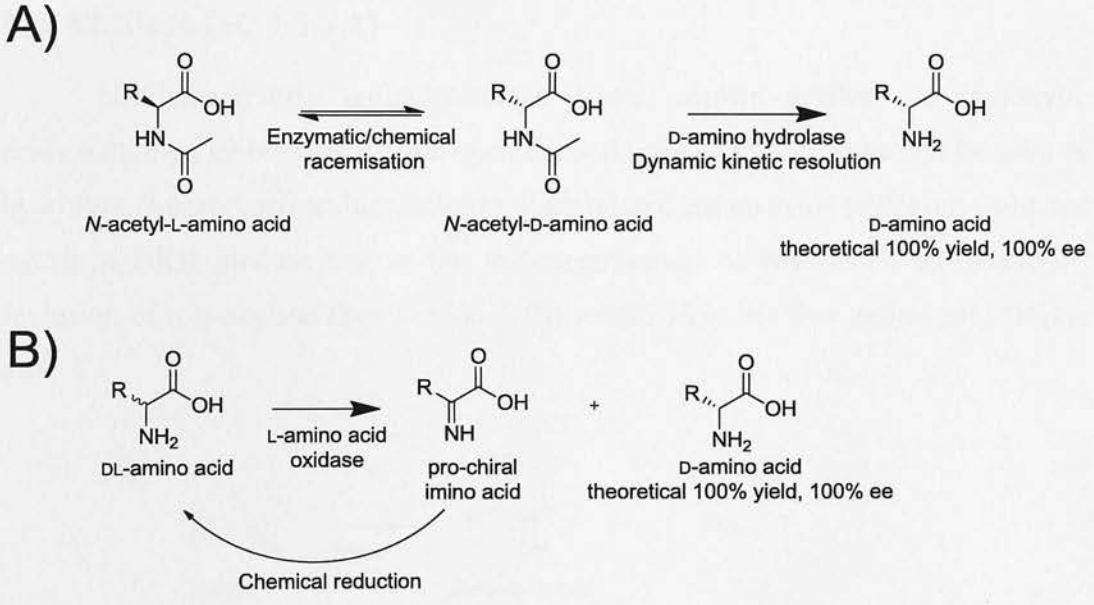


Figure 4. A) The dynamic kinetic resolution of a D-amino acid by *in situ* racemisation of an *N*-acetyl amino acid and hydrolysis by a D-amino hydrolase. B) De-symmetrisation of a chiral amino acid to generate a D-amino acid in 100% yield and e.e. with an L-amino acid oxidase and a chemical reduction step.

1.2.3, Asymmetric synthesis:

The final reaction involves the biotransformation of a prochiral starting material to give an enantiopure product. An example of this type of reaction would be the reductive amination of a keto acid by an oxidoreductase to give an enantiopure amino acid (**Figure 5**). Such reactions suffer from the need for an amine donor in stoichiometric amounts relative to the substrate which increases the overall cost.

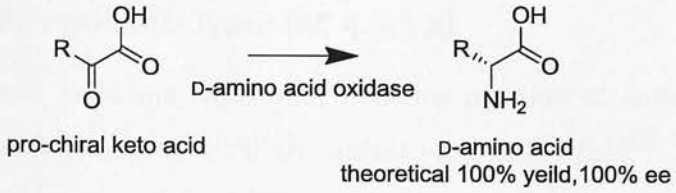


Figure 5. Asymmetric synthesis of a D-amino acid using a D-amino acid oxidase by reductive aminiation of a keto acid.

Often, any of these three possible reaction types can be used to generate an enantiopure amino acid, and the decision on which route to take inevitably comes down to substrate availability/cost, and of course the availability of a biocatalyst capable of the desired transformation. The list of possible biocatalysts is ever expanding; however there are several class of enzyme which make up the bulk of those that can be applied to the production of α-amino acids. A list of these classes and brief examples are given below.

1.3, Nitrilase (EC 3.5.5.1)

Nitrilases (Nitrile aminohydrolase, NIase) convert nitriles into carboxylic acids with high enantio- and regio-specificity (**Figure 6A**)[19, 20]. They can be used to hydrolyse *N*-acetylated aminonitriles to *N*-acetylated amino acids with high yield and e.e. in a DKR process due to the self-racemisation of *N*-acylated aminonitriles. Inclusion of a D-acylase (See section 1.10) would yield the free amino acid (**Figure 6B**)[21].

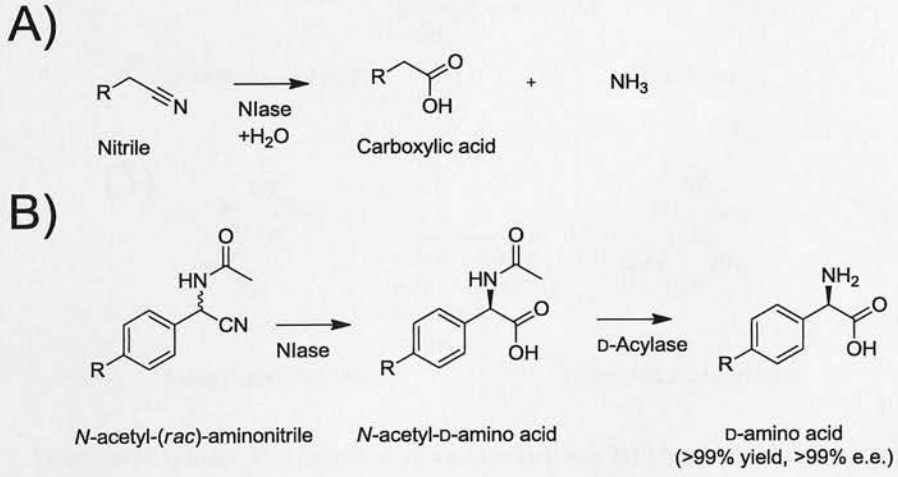


Figure 6. A) NIase hydrolysis of a nitrile to a carboxylic acid. B) NIase hydrolysis of an *N*-acetylated aryl-aminonitrile to yield an enantiopure *N*-acetyl-D-amino acid. This could then be de-protected by a D-acylase to yield the free amino acid.

1.4, Amino acid ammonia lyase (EC 4.3.1.x)

Amino acid ammonia lyases carry out the addition of ammonia to an α,β-unsaturated carboxylic acid to yield an enantio-pure amino acid[22, 23]. This class of enzyme was first employed in 1960 for the production of L-aspartic acid from fumaric acid by Japanese company Tanabe Seiyaku[24]. Common amino acid lysases include aspartic acid ammonia lyase (Aspartase), phenylalanine ammonia lyase (PAL), histidine ammonia lyase (HAL), and methyl-aspartic acid ammonia lyase (MAL) (**Figure 7**)[25-28].

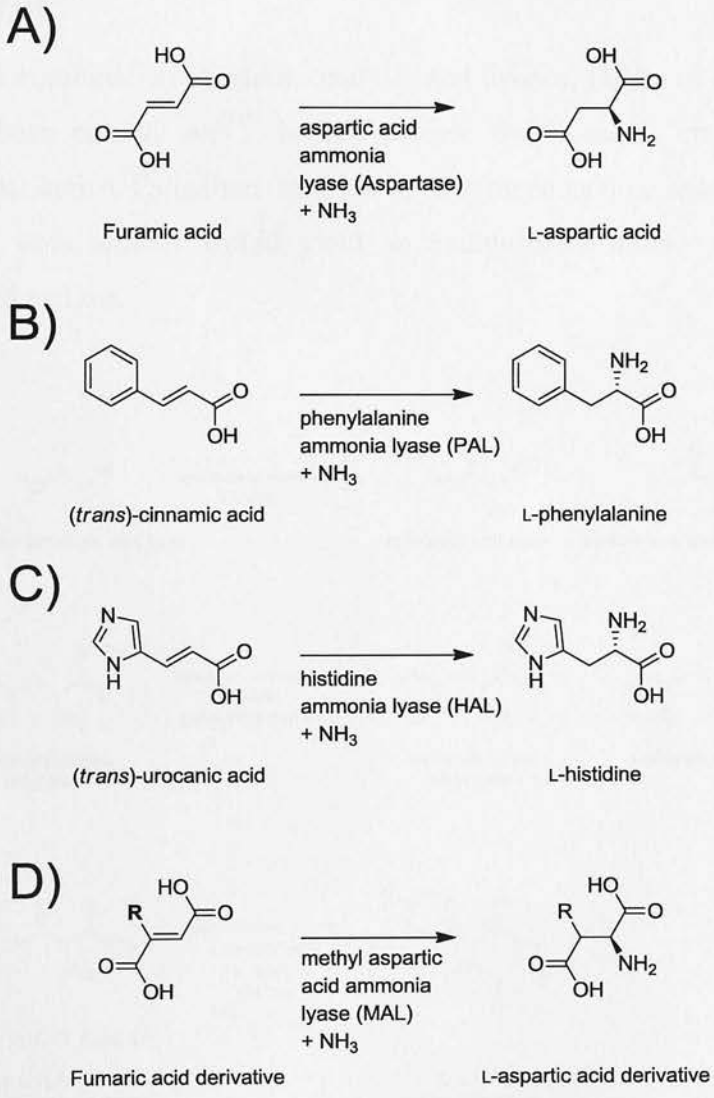


Figure 7. Amino acid lyases: A) aspartic acid ammonia lyase. B) phenylalanine ammonia lyase. C) histidine ammonia lyase. D) methyl-aspartic acid ammonia lyase.

1.5, Lipase (EC 3.1.1.3)

Lipases can be used for the kinetic resolution of carboxylic acids, secondary alcohols and primary amines[29]. They are one of the most versatile and well used

biocatalysts due to their high regio- and enantio- selectivity. These enzymes also have a high tolerance to organic solvent making them especially appealing for organic synthetic use[30]. The lipase reaction is reversible with both ester hydrolysis and trans-esterification reactions being catalysed (**Figure 8A**). The reaction equilibrium can be pushed towards either product by selection of the solvent[2]. Using lipases, amino acids can be resolved from DL-amino acids to give an amino acid ester and an α-amino acid (**Figure 8B**)[31, 32].

1.5.1, Dynamic kinetic resolution of amino acid derivatives using lipase and a chemocatalyst:

Using a combination of chemocatalysis and lipases, DKRs of DL-amino acid amides have been carried out[33]. In this process the lipase is employed as the resolving agent, and a Palladium catalyst as the racemisation agent (**Figure 8C**). Hydrolysis of both amides would yield an enantio-pure amino acid in >99% theoretical yield and e.e.

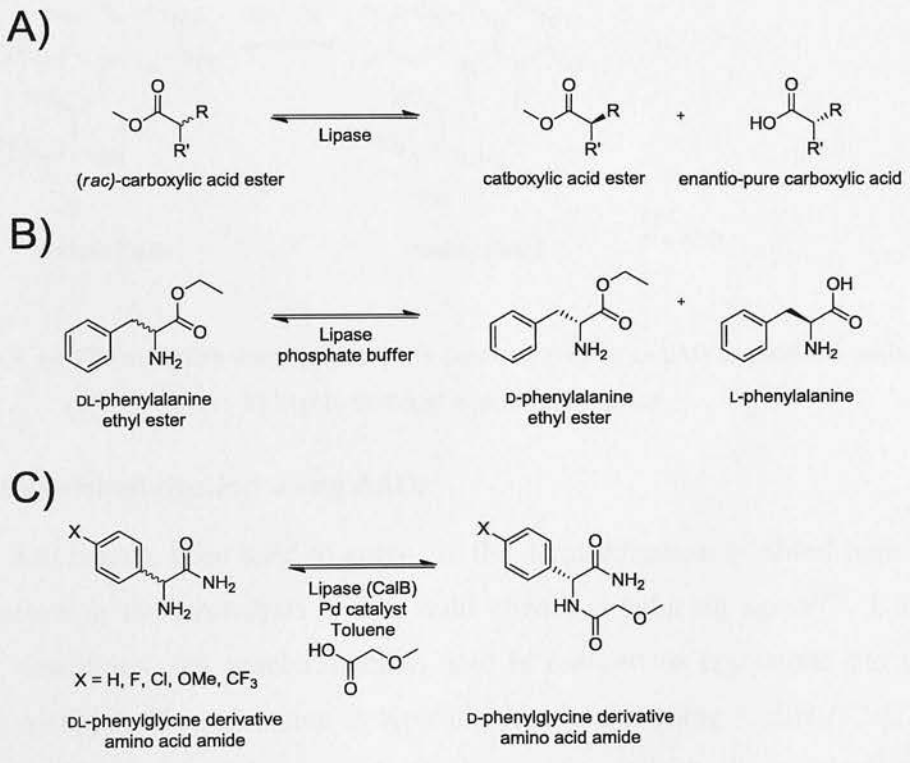


Figure 8. A) Kinetic resolution of an ester by hydrolysis. B) Kinetic resolution of L-phenylalanine from DL-phenylalanine ethyl ester. C) Combined chemo/biocatalytic DKR of enantiopure amino acid amides.

1.6, Amino acid oxidases (L-specific EC 1.4.3.2, D-specific EC 1.4.3.3)

Amino acid oxidases (AAOs) catalyse the stereoselective oxidative deamination of amino acids to their corresponding keto acid releasing ammonia as a side product (**Figure 9A**)[34-36]. Both L- and D- AAOs use flavin adenine dinucleotide (FAD) to shuttle electrons and hydrogen between the D/L-amino acid and molecular oxygen (**Figure 9B**)[37-39].

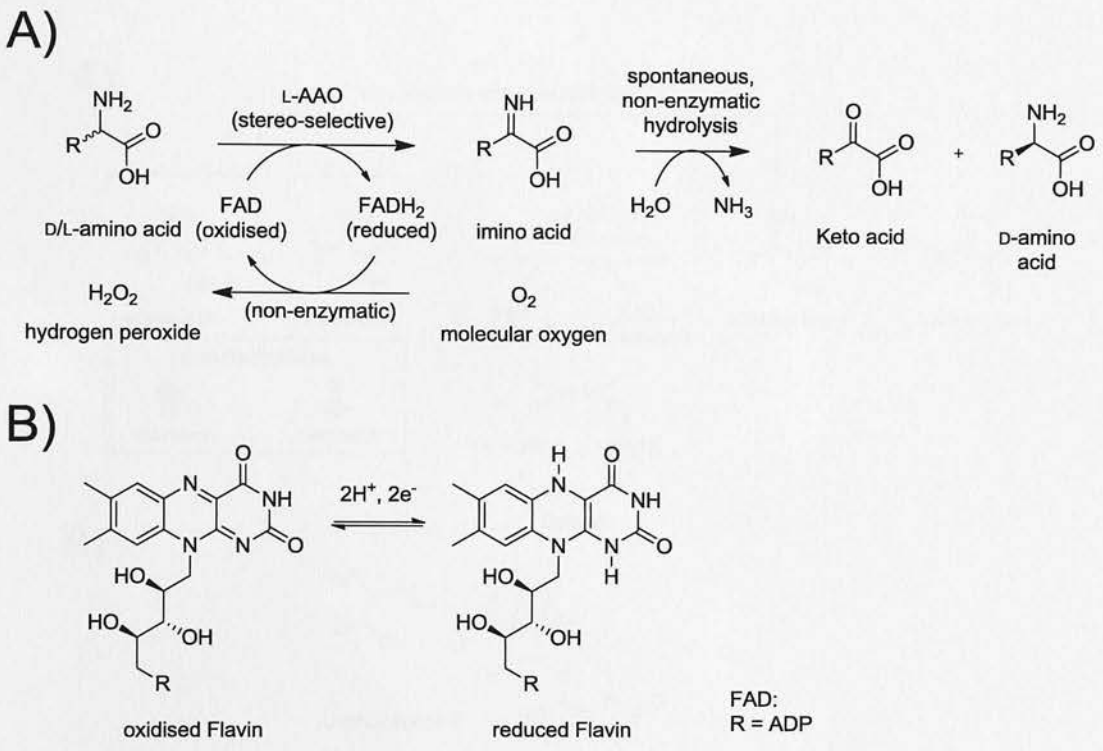


Figure 9. A) The oxidative deamination of an L-amino acid by L-AAO to yield a D-amino acid. B) Flavin oxidised and reduced forms.

1.6.1, Desymmetrisation using AAO:

AAOs have been used to carry out the deracemisation of chiral amino acids by interrupting the hydrolysis with a mild chemical reducing agent[40]. Under the correct conditions, the prochiral imino acid is reduced to regenerate the racemic starting material. This reduction is typically carried out using $NaBH_3(CN)$, amine-borane complexes or catalytic transfer hydrogenation, which allow reduction under conditions that are compatible with optimal enzyme activity (*i.e.* at neutral pH)[36]. This selective-oxidation/non-selective reduction gives rise to a cycle which will eventually leave only the enantiomer on which the AAO is un-reactive. Catalase (EC 1.1.1.6) can be incorporated into this reaction to remove the potent H_2O_2 oxidant, arising as a by-product of cofactor recycling (**Figure 10A**)[41].

1.6.2, Use of D-AAO in production of 7-aminocephalosporanic acid:

This D-AAO/Catalase system is employed in the industrial production of 7-aminocephalosporanic acid (7-ACA), a key intermediate in the synthesis of semi-synthetic penicillins, from cephalosporin C (**Figure 10B**)[39, 42-44]. A third enzyme, Glutaryl acylase (GAC, EC 3.5.1.93), hydrolyses the non-lactam amide bond in α-ketoadipyl-7-ACA, the D-AAO reaction product, to give the desired 7-ACA product[45].

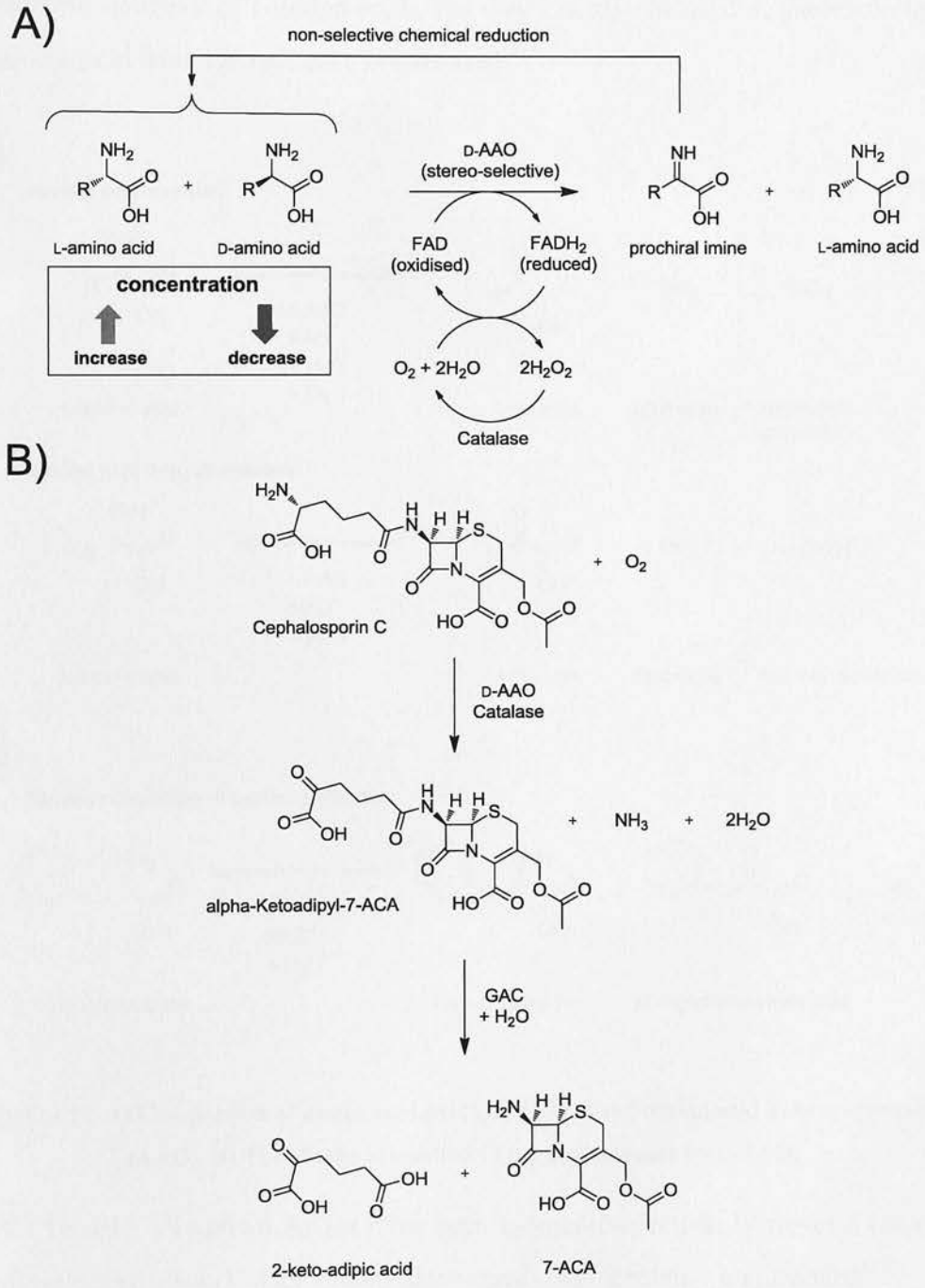


Figure 10. A) The deracemisation/ de-symmetrisation of amino acids using D-AAO. B) Production of 7-ACA from Cephalosporin C with D-AAO/Catalase/GAC.

1.7, Amino acid dehydrogenases (L-specific EC 1.4.1.5, D-specific EC 1.5.5.1)

Amino acid dehydrogenases (AADs) catalyse the oxidation of amino acids to their corresponding keto acids releasing ammonia as a side product in a similar fashion to AAOs (**Figure 11A**)[46]. The difference being that as ADDs use nicotinamide adenine dinucleotide (NAD^+), instead of the flavin cofactor used by AAOs, they do not generate hydrogen peroxide. ADDs are frequently used in the asymmetric synthesis of L-amino acids, but they can also be used in the resolution of a D-amino acid from the racemate (**Figure 11B**)[47].

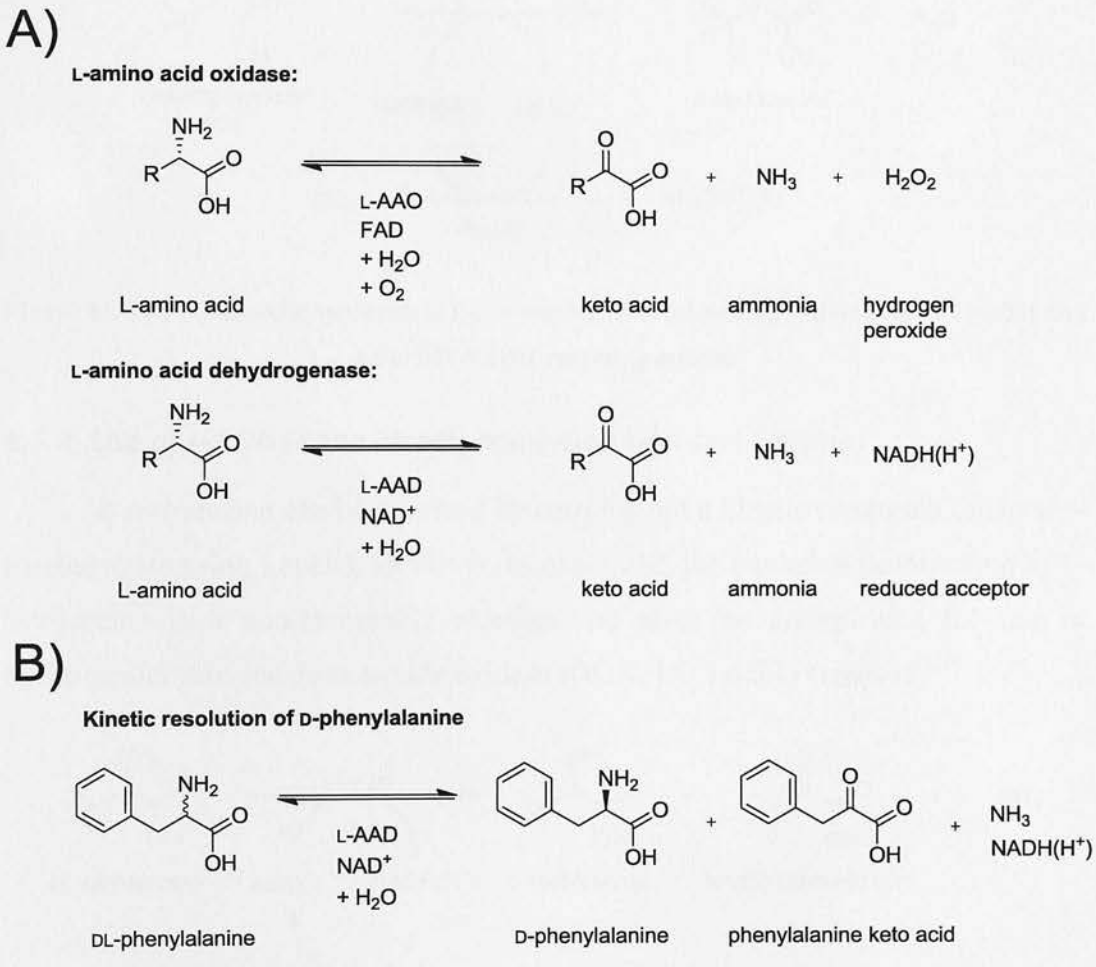


Figure 11. A) Comparison of amino acid oxidases (AAO) and amino acid dehydrogenases (AAD). B) The kinetic resolution of D-phenylalanine by L-AAD.

To date, all known AADs have been L-enantioselective; however a novel D-enantioselective AAD has been developed by protein engineering[48]. The dehydrogenase reaction is reversible and can be operated in reverse (synthesis direction) to perform asymmetric synthesis of enantiopure amino acids from

prochiral keto acids. The thermodynamics of this process favour the formation of the aminated product promoting high yields of the desired amino acid[46].

1.7.1, Use of L-ADD in the asymmetric synthesis of L-tert-leucine:

An established AAD process run in this manner is used to generate L-*tert*-leucine from the corresponding keto acid, trimethylpyruvic acid[49]. Leucine dehydrogenase (LeuDH, EC 1.4.1.9) is used in this process along with Formate dehydrogenase (ForDH, EC 1.2.1.43) which is used to recycle the expensive oxidised NAD^+ co-factor back to its reduced form (**Figure 12**).

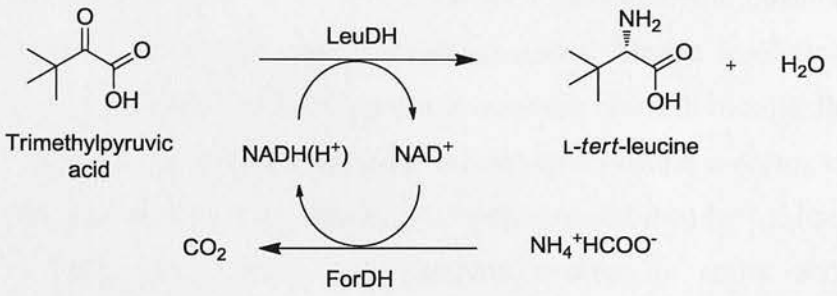


Figure 12. The asymmetric synthesis of L-*tert*-leucine from trimethylpyruvic acid by LeuDH and a ForDH NADH recycling system.

1.7.2, Use of L-ADD in the kinetic resolution of D-tert-leucine:

D-*tert*-leucine can be accessed by carrying out a kinetic resolution of DL-*tert*-leucine again using LeuDH. However, as explained, the oxidative deamination of L-*tert*-leucine is a non-favourable reaction and must be driven with the use of Nicotinamide adenine dinucleotide oxidase (NOX, EC 1.6.3.1) (**Figure 13**)[50].

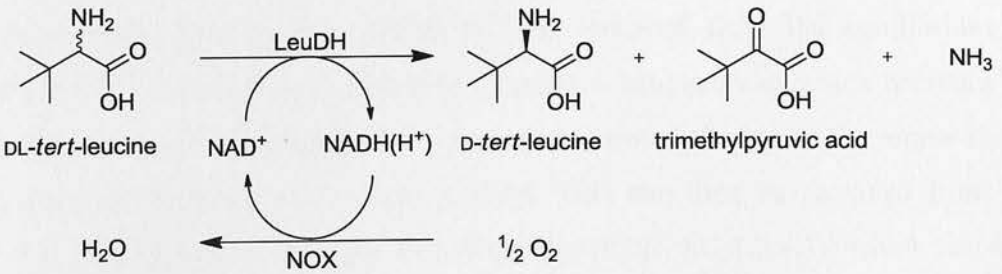


Figure 13. The kinetic resolution of D-*tert*-leucine from DL-*tert*-leucine with LeuDH.

1.8, Transaminases (EC 2.6.1.x)

Transaminases catalyse the transfer of an amino group from an amine donor to a keto acid in a pyridoxal-5'-phosphate (PLP) dependent reaction (**Figure 14A**)[51, 52]. The reverse reaction, using pyridoxal-5'-amine (PMP) as a cofactor, can be

referred to as an aminotransferase reaction. Transaminases can have both excellent regio- and enantio-selectivity, with both (*R*)- and (*S*)- transaminases being available[53]. Their regioselectivity has led to them being split into two types; α-Transaminases which are active on α-amine of amino acids and ω-Transaminases which are active on a non-α-amine in an amino acid, such as Orthinine transaminase (OTA, EC 2.6.1.13), or amino compounds which do not contain a carboxylic acid[54].

1.8.1, Kinetic resolutions using Transaminases:

The enantioselectivity of transaminases makes them ideal for the resolution of racemic amines; however, the need for at least stoichiometric quantities of the amine acceptor drives up the cost of these processes. Kinetic resolutions can be driven towards the product by use of an amine acceptor that will become the product for another enzyme, for example pyruvate can act as an amine acceptor to yield L-alanine which will then feed into the L-AAO pathway and then be oxidised back to pyruvate[55]. Using this method, only catalytic amounts of amine acceptor are required (**Figure 14B**).

1.8.2, Asymmetric synthesis using Transaminases:

When used in the synthesis direction (prochiral keto acid to amine product) enantiopure amines can be readily generated[56, 57]. This leads to a situation in which both enantiomers of an amine can be generated with a single enantioselective transaminase, either by kinetic resolution, or asymmetric synthesis[55]. Careful selection of the amino/keto donor can help to drive the synthesis direction reaction to completion, for instance the use of isopropylamine as an amino donor will generate acetone as the keto product which will be removed from the equilibrium by evaporation[18]. Another approach is to generate a keto product which becomes the substrate for a second enzyme, for instance, the use of alanine as the amine donor will generate pyruvate as the keto product. This can then be removed from the reaction equilibrium by Pyruvate dehydrogenase (PDH, EC 1.2.4.1)) which converts it to acetyl-CoA, however this approach introduces added complexity in the need for source of CoA (very expensive) and a Glucose dehydrogenase (GDH, EC 1.1.99.10) NAD^+ cofactor recycling system[53, 58].

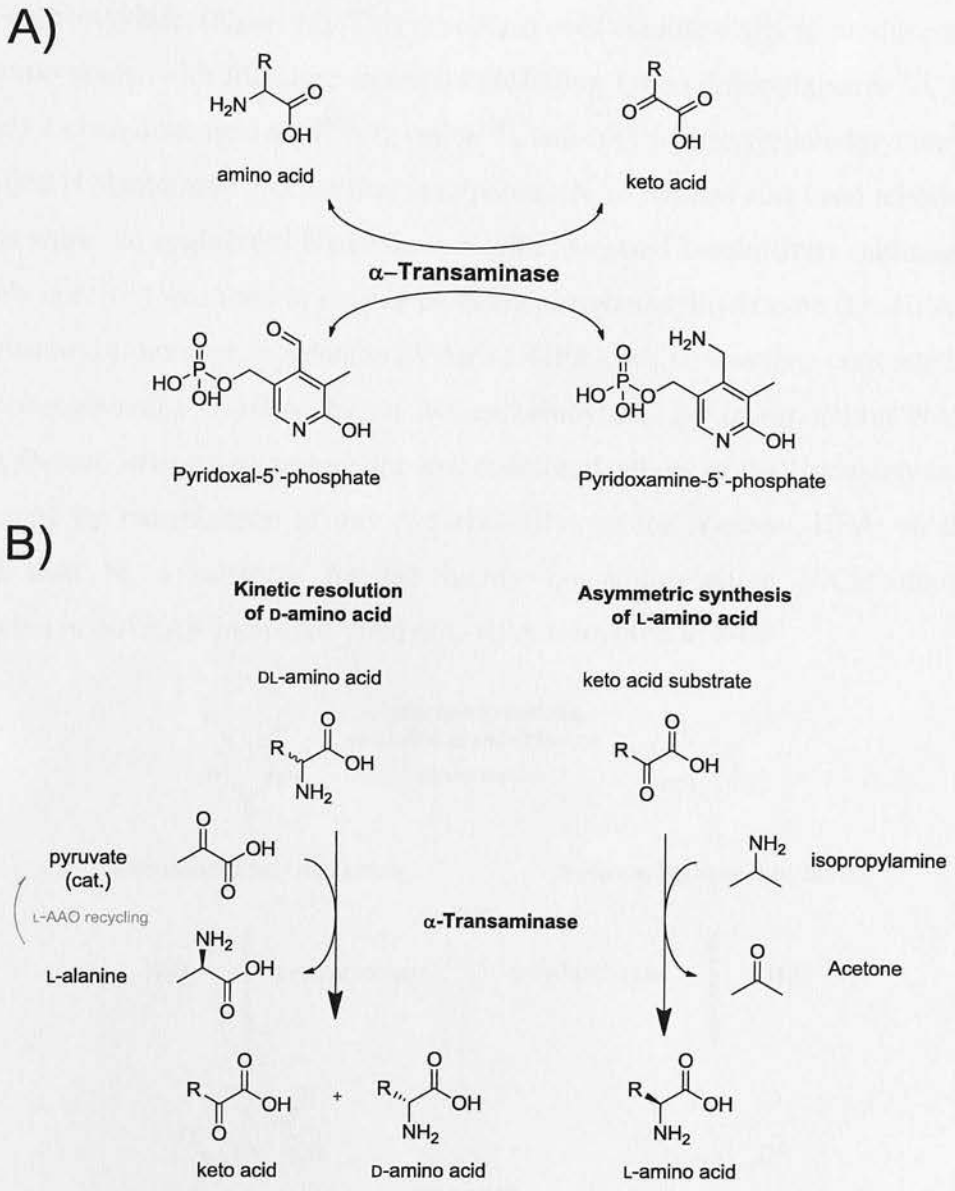


Figure 14, A) Transamination and aminotransferase reactions catalysed by an α-Transaminase and PLP/PMP. B) The kinetic resolution of a D-amino acid from DL-amino acid and asymmetric synthesis of the L-amino acid from the keto acid, both catalysed by the same L-selective α-Transaminase.

1.9, The Hydantoinase process

The Hydantoinase process utilises a Hydantoinase (Dihydropyrimidinase) (L-specific EC 3.5.2.2 / D-specific EC 3.5.2.3) and Carbamoylase (L-specific EC 3.5.1.77 / D-specific EC 3.5.1.87), to resolve a cyclic, racemic hydantoin into an enantiopure amino acid *via* a multi-step process (**Figure 15**)[59, 60]. The use of a Hydantoin racemase (EC 5.1.99.5) or suitable reaction conditions (> pH 8.0) allows this to be a DKR process[61]. The cyclic hydantoin is first stereoselectively hydrolysed to give the *N*-carbamoyl-amino acid, which is then hydrolysed again by

an *N*-carbamoylase, (**Figure 15**). This process is used commercially to produce L- and D- amino acids, with literature examples including L-homophenylalanine[62], L-(-2-amino)-4-phenylbutanoic acid[63], D-valine[64], and D-(4-hydroxy)phenylglycine[65]. A modified Hydantoinase process that incorporates NAAAR has also been reported[66]. In this work, an engineered Hydantoinase with increased L-selectivity (although not entirely specific) was used to resolve DL-homophenylalanylhydantoin (DL-HPAH) to *N*-carbamoyl-L-homophenylalanine (*N*-carb-L-HPA), which was then converted to L-homophenylalanine (L-HPA) by an *N*-L-carbamoylase. By incorporating NAAAR (from *Deinococcus radiodurans*), the low enantioselectivity of the Hydantoinase was overcome by racemisation of any *N*-carb-D-HPA to the *N*-carb-L-HPA, so that it could then be a substrate for the highly L-enantioselective *N*-Carbamoylase. Inclusion of NAAAR increased yield of L-HPA from 60% to 90%.

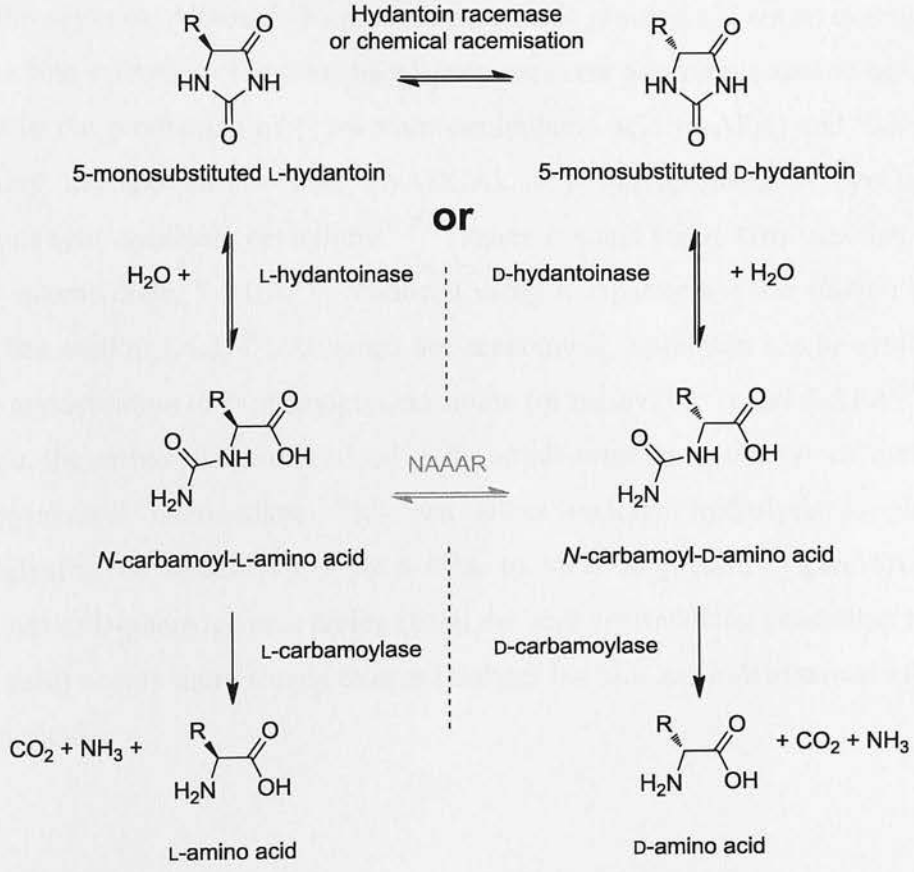


Figure 15. The hydantoinase process for resolution of both L- and D- amino acids.

1.10, Amidases (EC 3.5.x.x)

Amidases (also referred to as aminohydrolases or acylases) catalyse the hydrolysis of a amide bond to generate a amine and carboxylic acid, one of which is

usualy chiral (**Figure 16**)[67]. They are commonly used in the resolution of α-amino acids as both amino acid amides and *N*-acylated amino acids can be resolved.

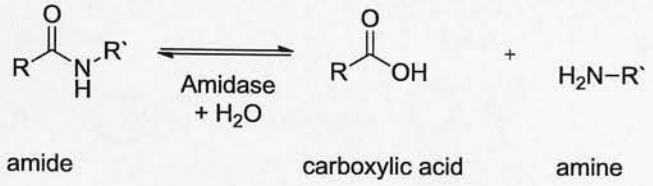


Figure 16. The amide hydrolysis reaction catalysed by an amidase.

1.10.1, The use of amidases (acylases) in the production of semi-synthetic penicillins:

Penicillin G and penicillin V amidases (PGA/PVA, EC 3.5.1.11) are two of the most commonly used industrial amidases[68-70]. They are commonly referred to as Penicillin acylases. Although the product from these processes is not an α-amino acid they are fine examples of acylase based processes and warrant discussion here. PGA is used in the production of (+)-6-aminopenicillanic acid (6-APA) and 7-amino-3-deacetoxy cephalosporanic acid (7-ADCA), key intermediates in synthesis of numerous semi-synthetic penicillins[71-73] (**Figure 17A** and **Figure 17B**). Another key β-lactam intermediate, 7-ACA, is produced using a separate acylase (GAC) and D-AAO (See section 1.4.2)[45]. Although not economical, ampicillin can be synthesised *via* the condensation of D-phenylglycine amide (or methyl ester) and 6-APA[74, 75]. In this case, the carboxylic acid is added as the amide/ester derivative which first forms an enzyme-acyl intermediate. This can either undergo hydrolysis to yield D-phenylglycine, or aminolysis with 6-APA to yield ampicillin (**Figure 17C**). The hydrolysis of D-phenylglycine amide (from the acyl intermediate, generating the free amino acid) occurs more slowly than aminolysis but still has a detrimental effect on product yield.

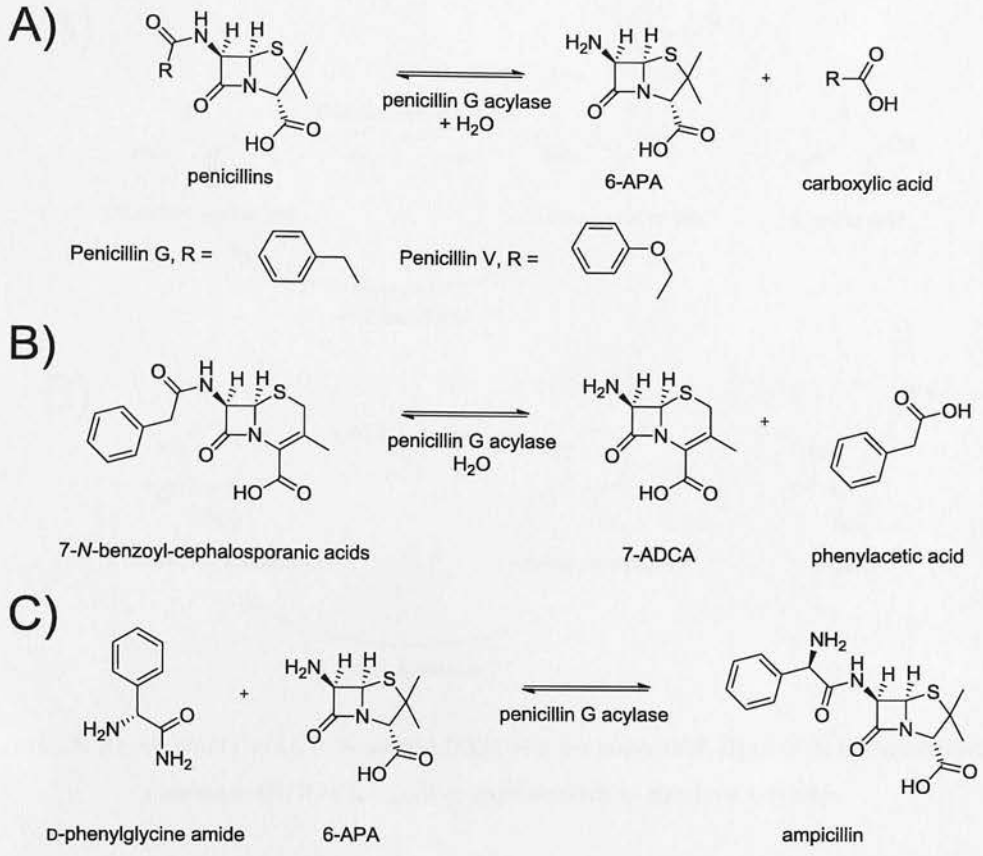


Figure 17. A) The acylase hydrolysis of penicillins to yield 6-APA B) The acylase hydrolysis of penicillins to yield 7-ADCA C) The acylase mediated synthesis of ampicillin *via* initial D-phenylglycine-enzyme (acyl-enzyme) intermediate formation, followed by condensation with 6-APA to yield the desired semi-synthetic ampicillin.

1.10.2, The use of amidases to resolve DL-amino acid amides:

Amidases can also be used to resolve DL-amino acid amides, in which case the carboxylic acid is the chiral product. The recent discovery that α-amino-ε-caprolactam racemase (ACL, EC 5.1.1.15) can racemise amino acid amides has led to the possibility of this process being implemented as a DKR (**Figure 18A**)[76, 77]. ACL racemase was originally used in combination with an L-ACL hydrolase to resolve DL-ACL to L-lysine (**Figure 18B**)[78].

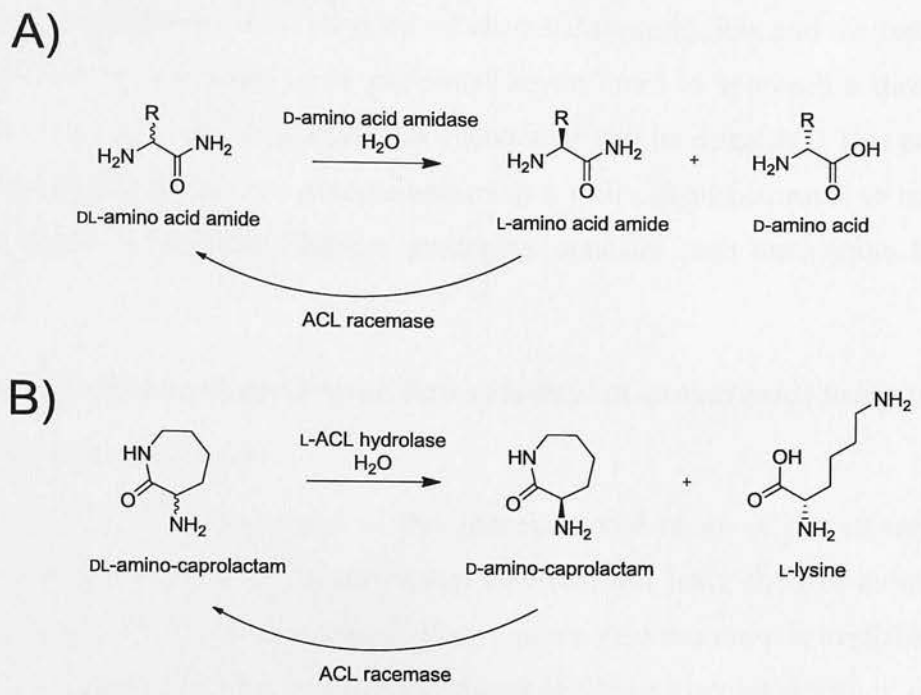


Figure 18. A) An amidase/ACL racemase DKR of a D-amino acid. B) L-ACL hydrolase/ACL racemase DKR of α-amino-ε-caprolactam to produce L-lysine.

1.10.3, The use of acylases in the kinetic resolutions of N-acyl-DL-amino acids:

Amidases (acylases) are key enzymes used in the resolution of *N*-acylated amino acids[79-81]. This process generates a waste carboxylic acid and chiral amine (amino acid)[82]. Their high activity and broad substrate tolerance, coupled with their excellent enantioselectivity makes them ideal biocatalysts. The original amidase enzymes were purified from mammalian sources such as porcine kidney, however many bacterial examples are known today, including an L-acylase from *Thermococcus litoralis* and D-acylase from *Alcaligenes denitrificans*, both of which are used industrially[83-88]. However, a flaw with these acylase based resolutions is that as the starting material is often chemically synthesised, it is racemic, resulting in 50% going to waste as the "off" enantiomer (**Figure 19**) [89].

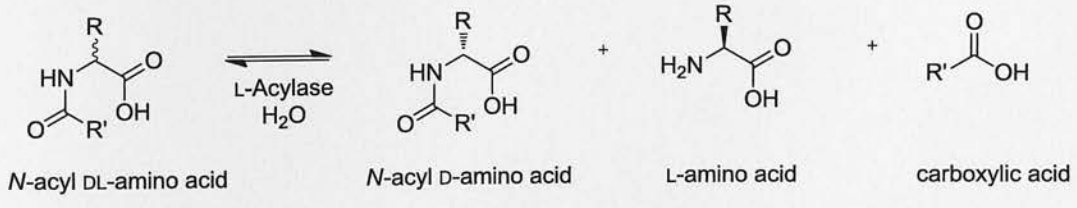


Figure 19. The resolution of an α-amino acid using an L-acylase.

The waste enantiomer can be recycled by chemical racemisation, and the resolution repeated, but as this needs to be performed seven times to approach a theoretical yield of 99%, it is more likely that this enantiomer will be discarded. This problem has not hindered the acylase process becoming a well-established route to many L- and D- amino acids, with Degussa producing optically pure methionine by this route[82].

1.10.4, The dynamic kinetic resolution of N-acyl-DL-amino acids using an acylase and a racemase:

An obvious improvement in this process would be an *in situ* racemisation step that would be specific for the *N*-acyl substrate and leave the free amino acid product in high e.e.[89]. A commercial *N*-acyl amino acid racemase is available (Rac 101, Biocatalytics), but this enzyme has several problems associated with it, such as poor activity towards the desired *N*-acetyl substrates and low thermostability. These observations are not surprising considering that Rac101 is a Proline racemase (ProR, EC 5.1.1.4) from *Clostridium sticklandii* (ATCC 12662), a non-thermophilic bacterium[90]. Many patents have been granted describing NAAAR enzymes and processes involving their use, but no enzyme has yet been disclosed that is effective economically under industrial conditions (for examples, see patents EP 1074628 B1, US 6372459, EP 0304021 B1, and US 7629154). If an industrially practical NAAAR could be developed it would allow acylase based processes to become even more useful and may allow for novel, cheaper routes that can compete with the established methods mentioned above.

1.11, Aims of this project:

The aim of this project is to investigate and develop possible industrial NAAAR biocatalysts that will allow the acylase kinetic resolutions of *N*-acetyl-DL-amino acids to become efficient DKRs (**Figure 20**).

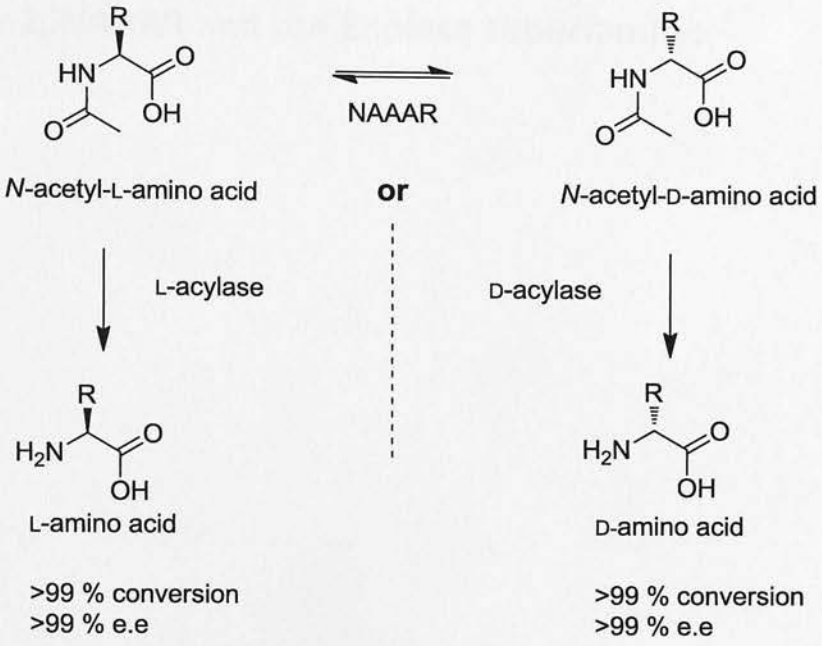


Figure 20. A NAAAR/acylase coupled dynamic kinetic resolution.

There are successful literature examples of NAAAR/acylase DKRs being utilised to produce enantiopure amino acids from *N*-acetylated racemic starting materials, but these also highlight the problems with low activity and substrate inhibition associated with NAAAR[91, 92]. The reported substrate inhibition above 50 mM would limit processes to less than 10 g L^{-1} for an average *N*-acetyl amino acid which is not cost effective and well short of the ideal 50 g L^{-1}[82]. To address these problems, a directed evolution approach was proposed in order to attempt to increase NAAAR activity towards a range of industrially relevant *N*-acetyl-amino acids at industrially relevant substrate concentrations.

Chapter 2, NAAAR and the Enolase superfamily:

2.1, Discovery of NAAAR:

NAAAR activity was first reported by Toukyama *et al* in *Streptomyces atratus Y-53* after screening of ~49,000 bacterial, actinomycetes, fungal and yeast strains[91, 93-95]. Further screening revealed a series of NAAAR enzymes from *Amycolatopsis sp.*, all of which share >90% homology with each other[96, 97]. Of these enzymes, NAAAR from *Amycolatopsis Ts-1-60* (*Am. Ts-1-60)* has the highest reported activity on *N*-acetyl derivatised substrates; however it is also reported to suffer from substrate inhibition above 50 mM. This inhibition is also described for a less active NAAAR homolog from *Amycolatopsis azurea,* but not in *Amycolatopsis orientalis subsp. lurida.* None of these enzymes have the required levels of activity that would justify the cost of their inclusion in an industrial process. Several other bacteria have been found to contain NAAAR enzymes, two of which, *Thermophilus thermus,* and *Deinococcus radiodurans*, have been crystallised, along with *Am. Ts-1-60*[98-100]. There are literature examples of NAAAR/acylase coupled DKRs being employed to produce enantiopure amino acids such as L-/D-methionine and L-homophenylalanine, however, there have been no published examples of this process applied at large scale[91, 92]. As a result, NAAAR is often neglected in literature reviews dealing with industrial enzymes, understandably replaced with more effective and commonly used biocatalysts. Despite this, *Am. Ts-1-60* NAAAR has been well studied by Gerlt *et al*, due to it being a member of the enolase superfamily[101-106]. NAAAR was found to be a promiscuous enzyme catalysing both the afore mentioned racemase reaction and the dehydration of 2-hydroxy-6-succinyl-2,4-cyclohexadiene carboxylate (2-HSCHC) to give *o*-succinylbenzoate (OSB). The *o*-succinylbenzoate synthase (OSBS, EC 4.2.1.113) reaction was found to be 10 fold faster than the *N*-acetyl racemase reaction, leading to the NAAAR protein being reassigned as an OSBS with low levels of promiscuous NAAAR activity. Furthermore, it was found that NAAAR exhibited high activity (on par with its OSBS activity) towards *N*-succinyl-amino acids, leading to the NAAAR protein being re-labelled again as a bi-functional enzyme with both OSBS and *N*-succinyl amino acid racemase (NSAAR) activity. In this work, the OSBS/NSAAR enzyme is referred to as an *N*-acetyl amino acid racemase (NAAAR).

2.2, Enolase superfamily:

Within the enolase superfamily, NAAAR is part of the muconate lactonising enzyme (MLE) sub-group[103]. This superfamily contains >1000 individual enzymes, all of which are structurally homologous, but display highly divergent sequences and mechanisms[107]. In terms of mechanism, each member of the enolase superfamily catalyses an initial proton abstraction from an α-carbon to generate an enediol intermediate, from which differing reactions can occur (**Figure 21**)[102]. This superfamily has been studied as the paradigm example of divergent enzyme evolution. It is thought that a progenitor gene, which had the ability to carry out the abstraction of an α-proton, has undergone gene duplication numerous times, and each time a novel advantageous function has been evolved for this new protein which required an initial proton abstraction. After many years of evolution, this process has given rise to the enolase superfamily we have today[108, 109].

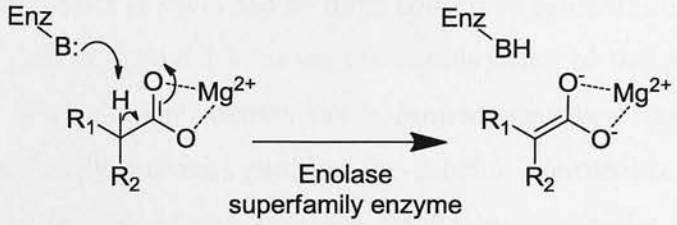


Figure 21. Initial proton abstraction step to form the enediol intermediate.

The substrates on which the enolase superfamily act is vast, ranging from amino acids and sugars to unsaturated carboxylic acids, however all substrates contain an essential carboxylate group to facilitate enediol formation[110]. The chemistry carried out by enolase enzymes is equally varied, with examples including lactone ring formation by Muconate lactonizing enzyme (MLE, EC 5.5.1.1), carbon-carbon double bond formation by enolase (EC 4.2.1.11) and β-ammonia elimination by β-methylaspartate ammonia lyase (MAL, EC 4.3.1.12))[111-113]. An overview of each enolase subgroup and the nature of the reactions contained within these groups can be seen below[114] (**Figure 23**). The initial reaction step of all enolase superfamily reactions forms a prochiral enediol intermediate by abstraction of the α-proton. The substrate on which this occurs, and the resulting chemistry that is carried out on the intermediate, is determined by the amino acid residues located on the *N*-terminal strands on seven of the eight β-sheets that make up the central $(\beta/\alpha)_7\beta$-barrel[115] (**Table 1**).

Table 1. Identities of amino acid residues located on the *N*-terminal ends of β_2- β_8 strands. Including Mandalate Racemase (MR, EC 5.1.2.2), dTDP-glucose 4,6-dehydratase (GlucD, EC 4.2.1.46), and Mannonate dehydratase (ManD, EC 4.2.1.8).

Sub-group	β_2	β_2	β_4	β_5	β_6	β_7	β_8
Enolase	E	D	E	D	K	H	K
MAL	H	D	E	D	K	-	-
MLE	KxK	D	E	D	K	-	E/DXD/G
MR	Kx(KRHDY)	D	E	E	D	H	E
GlucD	KxK	D	E	N	D	H	D
ManD	R	D	E	E	R	H	E

The enediol intermediate is stabilised by three conserved acidic residues that interact through a Mg^{2+} ion complexed between the carboxylates of the amino acids and substrate. Once this common intermediate is formed, structural variations between the enolase superfamily enzymes partition the enediol intermediate down differing reaction pathways to give various products. Mutations within the β-sheet terminal residues have been shown to interchange the activities of enolase enzymes. OSBS activity was imparted to both *E. coli* MLE II and *Pseudomonas* sp. *P51* L-Ala-D/L-Glu epimerase (AEE, EC 4.2.1.-) by the installation of single E323G and D297G mutations respectively[116-118]. These characteristics point towards the enolase superfamily active site having a degree of "plasticity", *i.e.* it is very susceptible to changes in substrate preference and reaction pathway through single mutations. This would make it suited to evolution, and especially divergent evolution where a single mutation in a duplicated gene can quickly generate a novel activity which provides a selective growth advantage. Several publications discuss the evolution of structure and function of α/β-barrel enzymes, with some advocating evolution from an ancient single progenitor and others claiming many progenitor genes have converged on this ideal protein fold (**Figure 22**)[114, 119, 120]. The α/β-barrel is clearly a stable protein configuration, as ~10% of all solved protein structures make use of it. The various cofactors employed by α/β-barrel enzymes (simple divalent cations, ribonucleotides and flavins), coupled with the vast range of functions catalysed within the barrel and the low sequence homology, would most likely suggest that convergent evolution has

driven separate progenitors towards this fold[121]. It is still unclear whether several ancestral enzymes fell into this thermodynamically favourable conformation independently of each other, or whether a single gene gave rise to every present day α/β-barrel gene. High homologies between the two halves of the α/β-barrels involved in histidine biosynthesis suggest that the eight strands of the barrel may have originated as two separate four stranded half barrels which have fused together[122]. The exact evolutionary history of this fold cannot be traced (proteins and DNA do not fossilise), and genomics/structural studies have not yet confirmed either scenario.

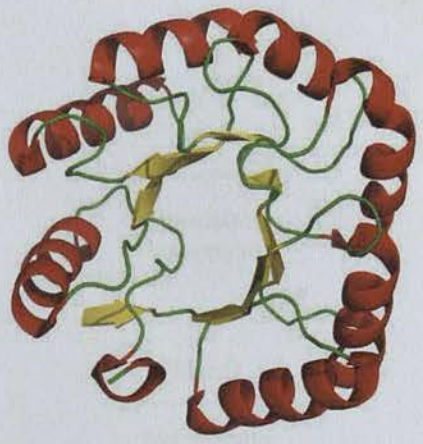


Figure 22. A typical β/α-barrel protein fold. Figure generated from published NAAAR: *N*-acetyl-L-methinoine structure, PDB:1SJA[100]. The active site would be found between the yellow β-strands.

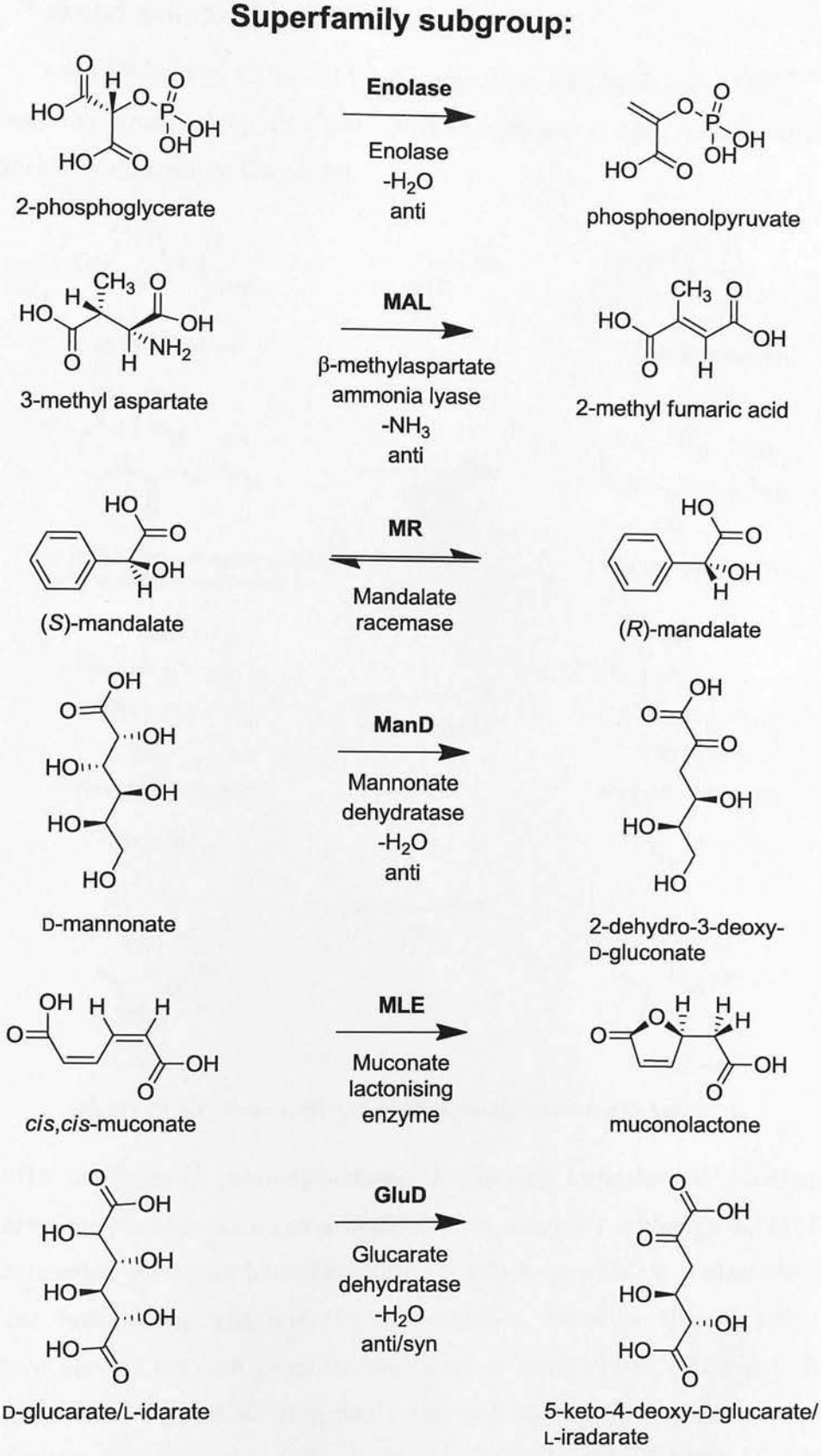


Figure 23. Examples of the paradigm reaction found within each subgroup.

2.3, *N*-acetyl amino acid racemase (NAAAR):

NAAAR belongs to the MLE subgroup along with MLE and AEE[103, 123]. Of interest, this small group of three enzymes catalyse a considerable variety of chemical transformations (**Figure 24**).

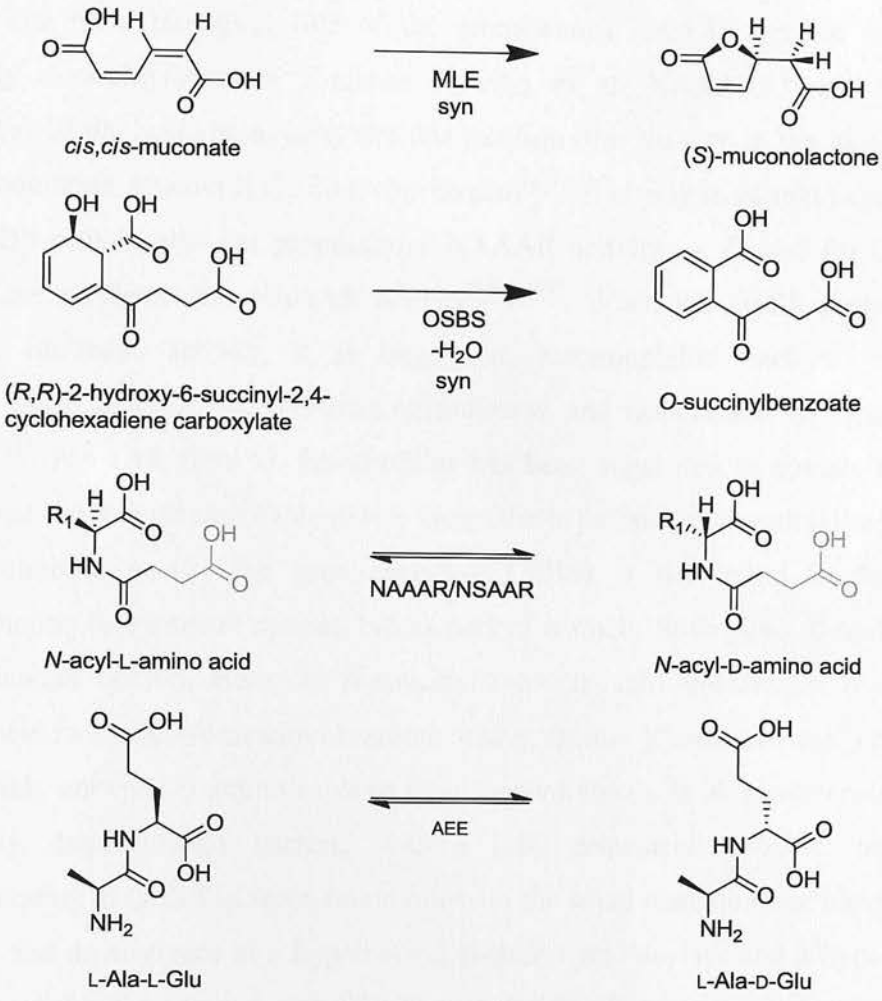


Figure 24. Reactions catalysed by MLE enolase superfamily subgroup.

The MLE subgroup of enzymes performs lactone ring formation, dehydration, and racemisation/epimerisation on structurally diverse substrates. Although the MLE and OSBS reaction substrates have very different structures, there is a clear similarity between those of the AEE and NAAAR reactions. However, there is only 22% sequence identity between putative *Amycolatopsis mediterranei U32* NAAAR and AEE (See section 6.22.2 for alignment). This is a common trait within the enolase superfamily; they are structurally similar but differ greatly in terms of catalytic mechanism and sequence. Whether NAAAR also has substantial AEE activity has not yet been determined, however low activity (<10% of that compared to racemase activity towards *N*-acetyl-L-methionine) has been reported with L-Ala-L-Met from

Am. Ts-1-60 NAAAR[95]. This would support the thinking that although these enzymes operate on related substrates *via* similar mechanisms, they have evolved separately to carry out different roles within the cell.

2.3.1, Biological role of NAAAR:

The main biological role of the promiscuous NAAAR protein is almost certainly *o*-succinylbenzoate synthase activity, as all NAAAR/OSBS enzymes, regardless of the host strain, carry out this reaction (the 3rd step in the biosynthesis of menaquinone, vitamin K12, from chorismate)[103]. It is only in certain bacteria that the OSBS protein also has promiscuous NAAAR activity; in *E. coli* for instance, OSBS has no detectable NAAAR activity[100, 103]. When the OSBS protein does possess racemase activity, it is largely in extremophillic bacteria, such as *Thermophilus thermus, Deinococcus radiodurans,* and *Geobacillus kaustophilus*[98, 99, 124, 125]. NAAAR from *G. kaustophilus* has been suggested to operate as a bi-functional enzyme playing a role in two biosynthetic pathways (moonlighting)[106]. In this bacterium, *menC* (the gene encoding OSBS) is not found in the usual menaquinone biosynthetic operon, but as part of a small, three gene, D-amino acid detoxification operon. Here, an *N*-succinyl-D-amino acid transferase/ *N*-succinyl amino acid racemase/ *N*-succinyl-L-amino acid hydrolase ("L-succinylase") pathway irreversibly converts D-amino acids to their L-enantiomers. In *A. mediterranei U32*, the only *Amycolatopsis* bacteria with a fully sequenced genome, the gene corresponding to OSBS is again found out-with the usual menaquinone biosynthesis operon, and downstream of a hypothetical D-amino acid acylase and a hypothetical member of the *N*-acyltransferase (NAT) superfamily. These two genes, along with NAAAR, may be part of an L-amino acid racemase pathway. Given that D-glutamate and D-alanine are essential components of peptidoglycan, NAAAR may play a role in the synthesis of these un-natural amino acids before they are incorporated into the peptidoglycan monomers to be transported to the periplasm[126]. The *G. kaustophilus* pathway appears to involve inversion of D- to L-, whereas the *A. mediterranei U32* performs this in the other direction, suggesting that in these cases NAAAR has been employed in two very different biological routes, a D-amino acid biosynthesis pathway, and a D-amino acid de-toxification pathway. Both of these require stereo-inversion to achieve different goals, so although the enzymes have evolved the same chemical activity, their roles within their host organisms may be different.

2.3.2, Substrate specificity:

NAAAR exhibits a large preference for *N*-succinyl-containing substrates (2-SHCHC and *N*-succinyl amino acids) over *N*-acetyl-containing substrates[101, 106, 124] **(Table 2)**.

Table 2. Published kinetic parameters for *Am. Ts-1-60* NAAAR.

Substrate	k_{cat} (s^{-1})	K_m (*mM*)	K_m/k_{cat} (M^{-1} s^{-1})
2-SHCHC	120	0.48	$2.5x10^5$
N-succinyl-L-methionine	110	5.9	$1.8x10^4$
N-succinyl-D-methionine	110	6.5	$1.7x10^4$
N-acetyl-L-methionine	6.4	11.0	$5.9x10^2$
N-acetyl-D-methionine	4.8	9.8	$4.9x10^2$

Upon examination of the *Am. Ts-1-60* NAAAR: *N*-succinyl-methionine complex X-ray crystal structure (PBD code: 1SJC) published by Gerlt *et al*, key interactions are observed that are thought to contribute to this substrate preference[100]. These are hydrogen bonding interactions between the G291 backbone carbonyl and the substrate amide, the S135 hydroxyl and substrate carbonyl, and finally, the R299 guanidinium group and the succinyl carboxylate *via* two structural H_2O molecules **(Figure 25)**.

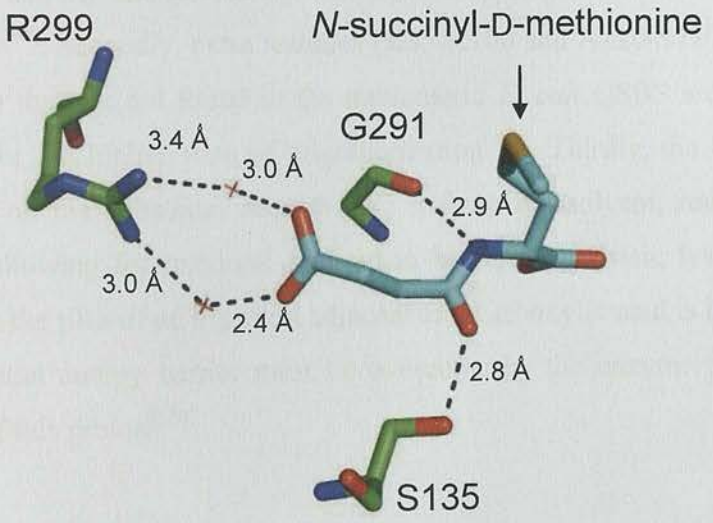


Figure 25. NAAAR enzyme interactions with *N*-succinyl-D-methionine highlighting the acyl-binding pocket. Red crosses represent H_2O molecules. Figure generated from 1SJC[100].

In contrast, all of these interactions are absent in the NAAAR: *N*-acetyl-L methionine complex structure (PDB code: 1SJA). In fact, apart from the common magnesium: substrate carboxylate interaction, there are no other obvious protein:ligand contacts that would explain substrate specificity. NAAAR generally shows a preference for L-enantiomers over D-, with activity being generally ~30% higher with the former substrate, although for certain amino acids, phenylalanine, leucine, and alanine this is reversed[95]. Low activity (50 times lower than the *N*-acetyl-methionine equivalent) has been reported with *N*-carbamoyl-methionine[124, 127]. NAAAR shows activity towards a wide range of amino acids, with highest activity exhibited towards those that are hydrophobic in nature. Activity appears to be especially low (100 fold less) with charged amino acids such as glutamate, aspartate, and arginine[106].

2.3.3, Protein structure:

Am.Ts-1-60 NAAAR is an octameric enzyme in solution, but hexameric and tetrameric isozymes have also been isolated from *Streptomyces atratus Y-53* and *G. kaustophilus* respectively[95, 125]. The *E. coli* NAAAR homologue, OSBS, is a monomeric enzyme[100, 128]. Like all enolase enzymes, NAAAR has a $(\beta/\alpha)_7\beta$-barrel fold, a derivative of the common $(\beta/\alpha)_8$-barrel protein fold, with each monomer containing two domains; an α+β *N/C*-terminal capping/lid domain (**Figure 26,** coloured red, M1-C132, G336-S368) and substrate binding $(\beta/\alpha)_7\beta$ barrel domain (coloured yellow, G133-A335). The domains are connected by a flexible hinge which allows the enzyme to close around the substrate. The capping domian has three roles in activity and structure. Firstly, it is responsible for amino acid substrate specificity[99, 105]. Secondly, extra residues (K85-K100 and A120-S129) found within the lid region that are not found in the monomeric *E. coli* OSBS are thought to be responsible for the higher state of oligomerization[100]. Thirdly, the capping region closes down on the substrate, sequestering it from the solvent, removing solvent effects and allowing for maximal interaction between the basic lysine and the α-proton[98]. As the pKa of an α-proton adjacent to a carboxylic acid is between 22 and 25, a substantial energy barrier must be overcome by the enzyme to facilitate the abstraction of this proton[129].

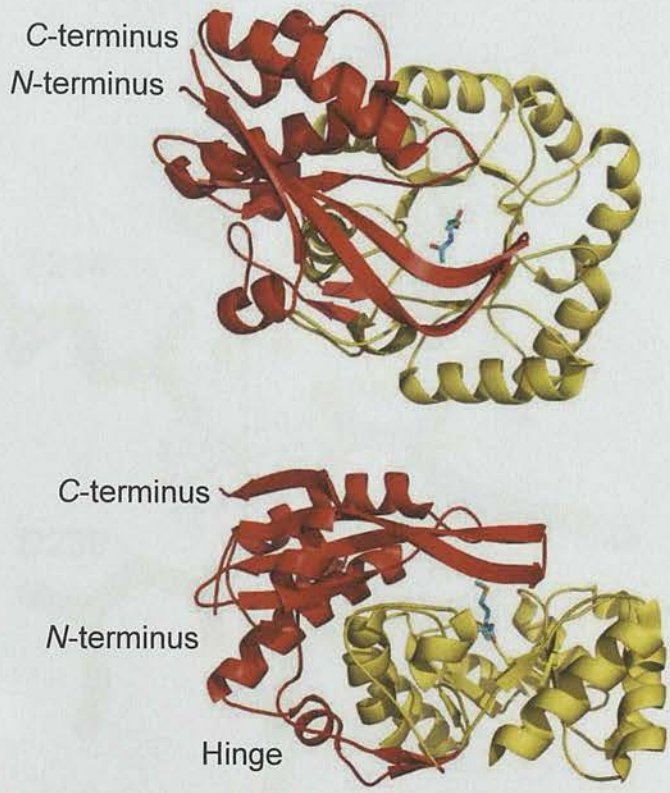


Figure 26. The *Am.Ts-1-60* NAAAR monomer, with the α+β *N/C*-terminal lid/capping region in red and $(\beta/\alpha)_7\beta$ barrel in yellow. The substrate is shown in cyan. Figure generated from 1SJA[100].

2.3.4, Reaction mechanism:

Three acidic amino acids (D189, E214, and D239) and a H_2O molecule bind the substrate *via* a Mg^{2+}, and two lysines (K163 and K263) carry out the proton abstraction/re-protonation responsible for racemisation[100, 101] (**Figure 27** and **Figure 28**).

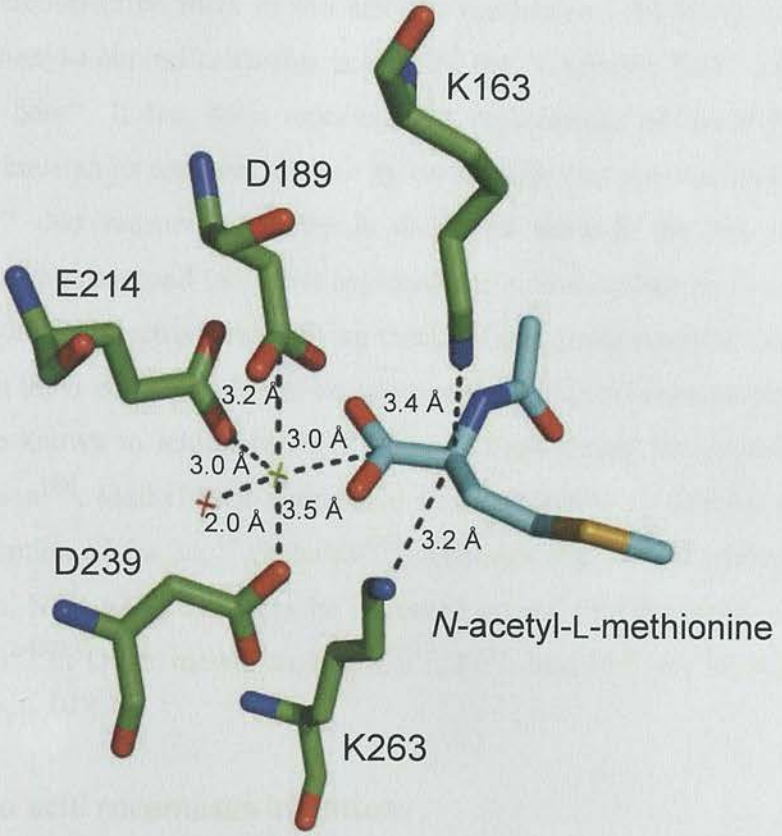


Figure 27. The interactions between *Am. Ts-1-60* NAAAR residues D189, E214, and D239 and Mg^{2+} (green cross) and H_2O (red cross). Catalytic K163 and K263 residues and *N*-acetyl-L-methionine are also shown. Figure generated from 1SJA[100].

In NAAAR, as with all enolase enzymes, the initial step is proton abstraction to generate an enediol intermediate, however as there is no further chemistry possible on the *N*-acyl enediol intermediate (for example loss of water or ammonia), the reaction effectively stalls and goes in reverse to regenerate the initial amino acid substrate. If the abstraction and subsequent re-protonation are carried out by differing lysine residues, then stereo-inversion around the α-carbon occurs, allowing for racemisation of the *N*-acyl amino acid substrate (**Figure 28**).

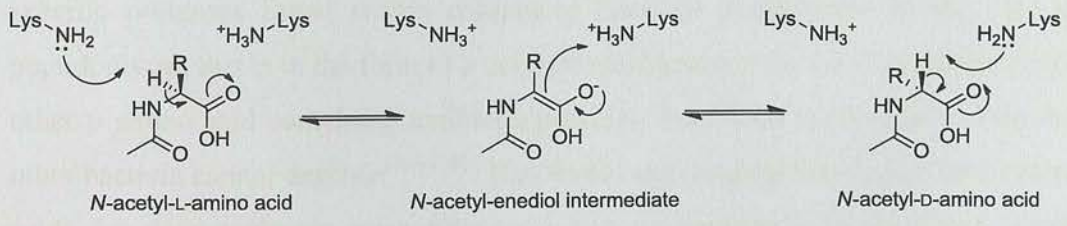


Figure 28. The mechanism of NAAAR-catalysed racemisation of *N*-acyl-amino acids.

The arrangement of these two lysine residues is such that they sit on either side of the substrate α-carbon allowing for K163 and K263 (numbered from *Am. Ts-1-60*) to be

assigned stereoselective roles in the enzyme mechanism. As K163 can only de-protonate *N*-acyl-L-amino acids this is labelled the "L-specific base" and K263 the "D-specific base". It has been reported that replacement of these lysines with arginine/serine/alanine removes all activity confirming they are essential for enzyme turn-over[101]. No racemase activity is displayed towards the free amino acid (essential for the proposed industrial application), indicating that an *N*-acyl group is crucial and that the electron withdrawing nature of this group promotes abstraction of the proton at the α-carbon[96]. Metal chelators such as ethylenediaminetetraacetic acid (EDTA) are known to inhibit NAAAR activity, highlighting the requirement for a divalent cation[93]. Methyl ester amino acid substrates show no activity, again likely due to disruption of the Mg^{2+} chelation[95]. Although Mg^{2+} is the natural cation for this reaction, NAAAR activity can be increased several fold by substituting this for Co^{2+} or Mn^{2+}[97]. Other metals such as Cu^{2+}, Fe^{2+}, and Pb^{2+} are known to inhibit enzyme activity[125].

2.4, Amino acid racemases in nature:

2.4.1, Role of D-amino acids in nature:

Although cell biology is based almost exclusively on L-amino acid chemistry, D-amino acids do serve a purpose in bacterial cells, for example, D-alanine and D-glutamate form essential components of bacterial peptidoglycan, and D-phenylalanine is incorporated in the peptide antibiotic Gramicidin S[130, 131]. Given that there are no known *de novo* D-amino acid synthesis pathways in nature, the need for amino acid racemases is obvious[132]. The use of D-amino acids has likely arisen after evolution of the L-amino acid biosynthetic pathways, because if this was not the case, the equivalent D-amino acids pathways would also surely exist. It is possible that the use of D-amino acids has evolved to impart resistance to L-amino acid specific proteases found within competing bacterial populations. In the case of peptidoglycan this is in the form of a defence mechanism, whereas Gramicidin S, and other D-amino acid containing antibiotic peptides, have been evolved as a toxin that other bacteria cannot degrade[133-135]. This would also suggest that the use of D-amino acids for defence/attack arose after evolution of L-amino acid chemistry within bacteria. As many amino acid synthetic pathways are multi-step and inter-connected, nature has decided to use a racemase to synthesise the D-enantiomer, following synthesis of the L-enantiomer[136]. Bacteria that have made this evolutionary step and

use this novel D-amino acid building block, give themselves an evolutionary advantage over their competing neighbours.

2.4.2, Mechanisms of amino acid racemases:

There are two mechanisms for amino acid racemisation that have been evolved by bacteria, the first of which generates a prochiral planar sp^2 hybridised carbon centre *via* an enolate, and the second, which generates an sp^3 hybridised carbanion. The most widespread mechanism involves the use of a PLP cofactor to generate this sp^3 hybridised carbanion. Examples of four bacterial racemases which use these mechanisms are given below.

2.4.1.1, Alanine racemase:

Alanine racemase (Alr, EC 5.1.1.1) from *Bacillus stearothermophilus*, a well-known and studied amino acid racemase uses PLP to generate a carbanion that allows racemisation (**Figure 29**)[137]. Alr initiates racemisation by formation of a substrate external aldimine with a transaldimination between the α-amino group of the L-alanine substrate, and the ε-amino group of K39 bound to PLP. A concerted proton abstraction by Y285 from the α-carbon and donation to the substrate carboxylate generates the carbanion. This is followed by a further proton abstraction by K39 from the substrate carboxylic acid, followed by donation of a proton to the carbanion. The location of Y285 and K39 on opposite sides of the α-carbon permit this process to carry out racemisation of the substrate. K39 then reforms the PLP-internal aldimine *via* another transaldimination and releases the free amino acid product, D-alanine.

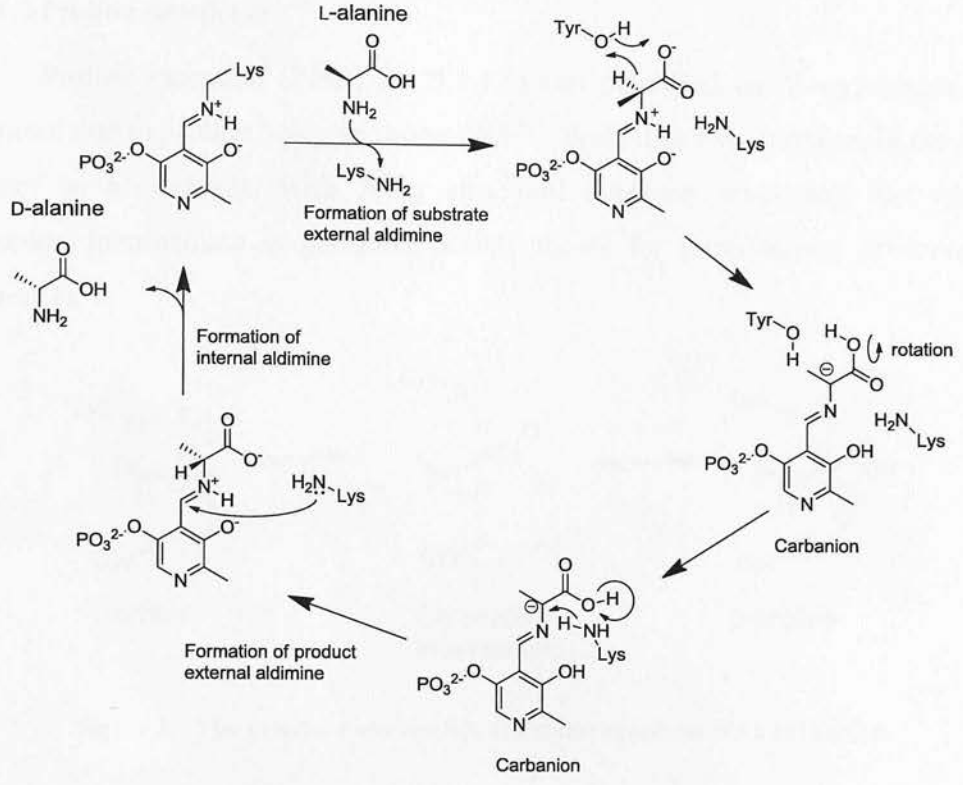


Figure 29. The proposed mechanism of the PLP dependent Alanine racemase.

2.4.1.2 Glutamate/Aspartate racemase:

Glutamate racemase (MurI, EC 5.1.1.3) uses a cofactor-free mechanism to generate a carbanion, with a pair of cysteines performing the acid/base chemistry[138, 139]. This mechanism is also found in Aspartate racemase (AspR, EC 5.1.1.13)[140]. Two cysteine residues act as stereospecific acid/bases to first abstract and then re-protonate the α-proton of the glutamate or aspartate substrate. X-ray crystal structure data would suggest that MurI operates similar to Alr, with the racemisation proceeding *via* the formation of a planar carbanion containing an sp^3 hybridised α-carbon (**Figure 30**)[141, 142]. Although there have been mutational studies, and the crystal structure solved, the exact roles of other amino acids within the MurI/AspR active site have not yet been fully elucidated.

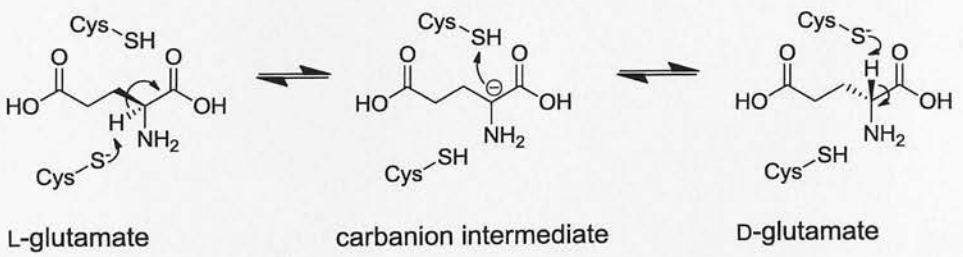


Figure 30. The proposed mechanism of Glutamate racemase *via* a carbanion.

2.4.1.3 Proline racemase:

Proline racemase (ProR, EC 5.1.1.4) can be called an *N*-acyl amino acid racemase due to proline being an imino acid[90]. ProR uses two cysteines in the same manner as MurI/AspR, with X-ray structural evidence suggesting that an sp^3 carbanion intermediate is generated which allows for stereo-centre inversion[143] (**Figure 31**).

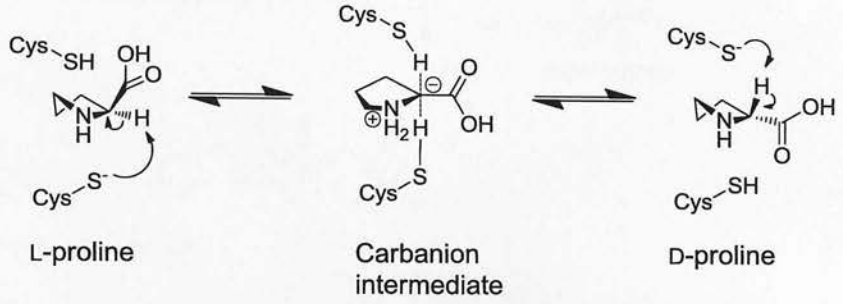


Figure 31. The proposed mechanism of Proline racemase *via* a carbanion.

2.4.1.4 Phenylalanine racemase:

A unique ATP-dependent Phenylalanine racemase (GrsA, EC 5.1.1.11) has also been discovered from *Bacillus brevis* in which racemisation proceeds *via* a thioester enolate[144, 145]. As opposed to Alr, MurI/AspR, and ProR which are individual enzymes, GrsA is part of gramicidin S synthetase, a multi-subunit and domain non-ribosomal peptide synthase (NRPS). Gramicidin S is an antibiotic containing two pentapeptides (valine-orthinine-leucine-D-phenylalanine-proline) joined in a cyclic form *via* two valine-proline peptide bonds. The gramicidin S synthase is comprised of two subunits, GrsA responsible for racemisation and peptide synthesis initiation with D-phenylalanine, and GrsB which then carries out the peptide elongation. Phenylalanine racemisation is carried out by GrsA in an ATP dependent manner *via* three domains; an adenylation domain, a thiolation domain, and an epimerisation domain (**Figure 32**).

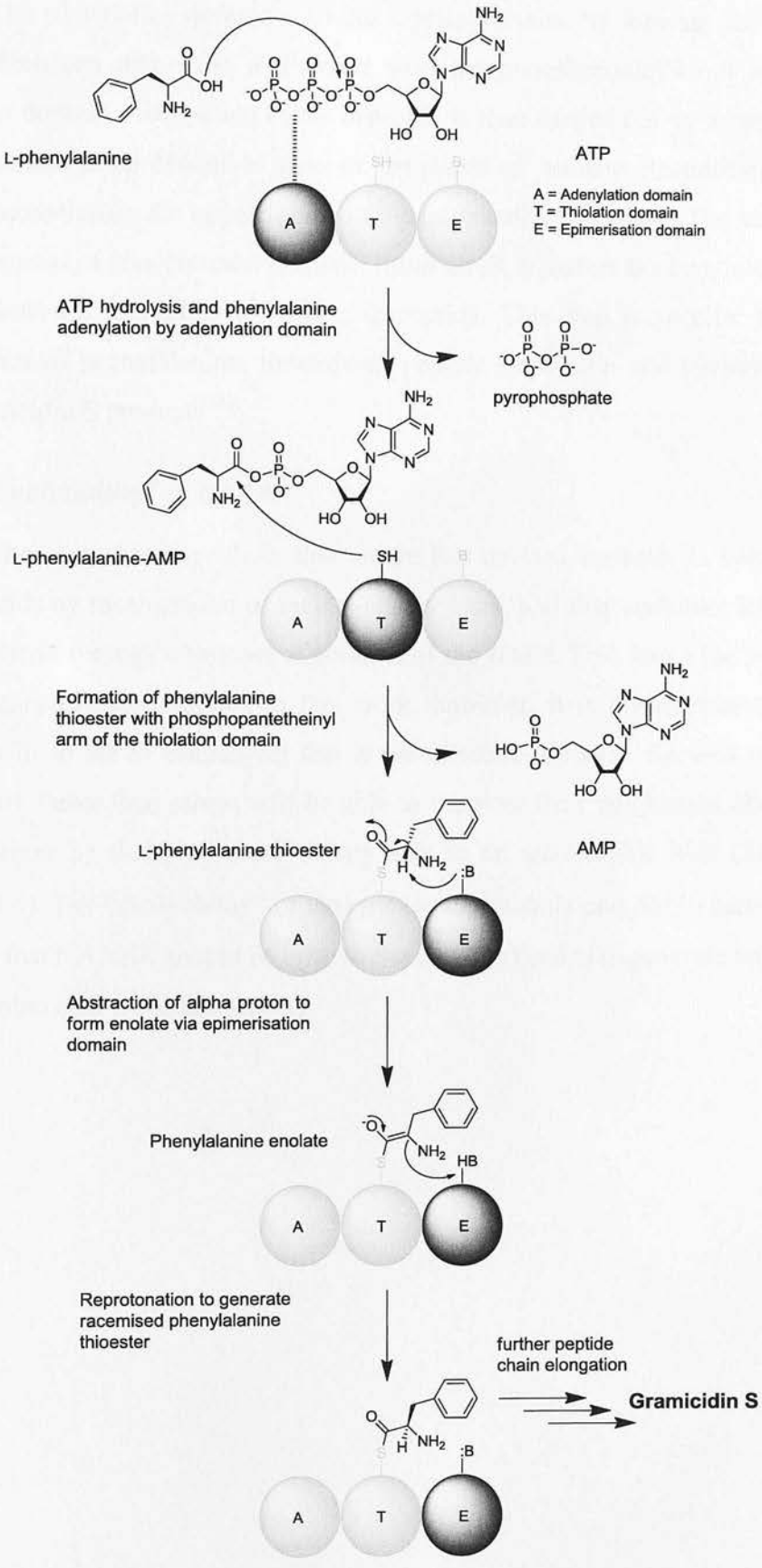


Figure 32. The proposed mechanism of ATP-dependent racemisation of L-phenylalanine by GrsA *via* a thioester enolate.

The adenylation domain activates L-phenylalanine by forming the Phe-AMP ester, which can then form a thioester with the phosphopantetheinyl arm of the thiolation domain. Abstraction of the α-proton is then carried out by a basic residue on the epimerisation domain to generate the planar sp^2 enolate. Racemisation occurs by re-protonation on the opposite side of the intermediate α-carbon. The next subunit in this process, a condensation domain within GrsB, transfers D-phenylalanine from the thioester on to proline to form a di-peptide. This step is specific for the D-enantiomer of phenylalanine. Subsequent peptide elongation and cyclisation gives the Gramicidin S product[146].

2.4.3, "Evolvability" of NAAAR:

These mechanisms show that nature has devised methods to synthesise D-amino acids by racemisation of their L-enantiomers, and that evolution has selected these bacteria through a process of survival of the fittest. This forms the basis of the evolutionary pressure utilised in this work, however, it is the racemisation of D-amino acids to the L- enantiomer that is the selective pressure. Bacteria that evolve this ability faster than others will be able to outgrow their neighbours allowing for identification by their increased colony size in an auxotrophic host (SET21, see section 3.6). The "evolvability" of the enolase superfamily and $(\beta/\alpha)_8$ barrel proteins suggests that NAAAR should be open to the changes needed to generate increased *N*-acetyl amino acid racemase activity.

Chapter 3, Directed evolution of enzymes:

3.1, Directed evolution:

At first, enzymes appear to be ideal catalysts for many chemical synthetic steps due to their excellent chemo-, regio- and enantio- selectivity[53]. However, the conditions under which a particular enzyme operates, and the substrates which it accepts, are not always compatible with those a chemist wishes to use, leading to enzymes often being far from ideal biocatalytic tools. Recently methodologies have been developed that allow for novel enzymes with improved catalytic properties to be generated. This process of mutagenesis and screening of generated variants in hope of discovering mutations that cause improved properties has been termed "Directed Evolution".

For an enzyme to form part of a successful biotransformation process, it must catalyse the reaction fast enough to give high yields of product in a reasonable time frame, possess the correct reaction specificity, be stable under the bioprocess conditions and be amenable to production at high volumes (fermentation) and subsequent downstream processing[147]. The enzyme may have to function under conditions that facilitate substrate/product stability, for instance, non-native temperature (<30°C or >37°C), non-neutral pH or in reaction media containing organic solvent[148, 149]. These requirements can cause problems as enzymes tend to display their highest activities on a very narrow substrate range and are typically only active in aqueous media at physiological pH and temperature[150, 151]. This situation has arisen because evolution has often selected their design towards immense specificity at the cost of multiple activities and broad substrate range[152]. The preference for mild reaction conditions is due to evolutionary pressure having been exerted under physiological conditions, as opposed to 50°C in 25% MeOH. As a result of these selective requirements, there are, with possibly only a few exceptions, such as lipases, no enzymes that can be used to carry out general reactions on a wide range of substrates under harsh chemical conditions[29]. The highly selective nature of enzymes has long been known and was described as being like a "key and lock" by Emil Fischer in 1894[153]. This is especially true when considering enzymes in biocatalytic reactions, as often a high number of keys must be tested to find one that fits the desired lock. The highly specific nature of enzymes is one of the major considerations when evaluating enzymes for use in synthetic reactions. When a reaction is being tested with a natural substrate or a molecule closely related to a natural substrate, screening can generally be performed directly with this compound.

For instance, screening for a leucine amino acid racemase operating at 37°C would involve testing this reaction using L-leucine and a variety of amino acid racemases. Once a positive hit has been discovered, if any characteristic of the enzyme is deemed unsuitable, attempts could then be made to tailor the enzyme using a directed evolution approach[154]. In more complicated scenarios, when there is less chance of a natural enzyme being compatible with a synthetic substrate or reaction conditions, such as screening for a *tert*-leucine amino acid racemase operating at 60°C, it may be necessary to initially screen under mild conditions using a natural substrate (L-leucine) that is structurally similar to the desired un-natural substrate, before optimising activity towards *tert*-leucine using directed evolution. In this way it may be possible to engineer thermostability into a protein which displays the desired specificity. Another approach would be to begin with thermostable amino acid racemases and screen these enzymes for activity towards *tert*-leucine/leucine. As an example of this approach, a novel Cytochrome P450 propane monooxygenase was engineered from a Long chain fatty acid (C_{12}-C_{18}) monooxygenase (EC 1.14.15.3) from *Bacillus megaterium*[155, 156]. Initial screening and directed evolution used an intermediate octane substrate, as the wild-type (WT) enzyme had no measurable activity towards the desired propane substrate. Octane was utilised for five rounds of screening, after which a novel propane active variant was generated. Eight further rounds of screening with propane increased the k_{cat} with the C_3 substrate from 0 to 450 min^{-1}. Using directed evolution, enzymes that are close to the requirements for industrial biocatalysts can be improved upon to generate novel processes that can compete or even replace the present day chemocatalytic routes[17].

3.2, Generating variant libraries:

Successful directed evolution generally requires mutagenic libraries containing thousands to millions of randomly mutated genes to be synthesised and screened effectively[16, 157]. Much smaller libraries may also be used where mutations are limited to a small number of amino acid positions in the target enzyme. The level of screening is therefore influenced by how well the mutagenesis can been rationalised, with minimal screening required for more focused libraries in which key catalytic, or substrate binding residues have been targeted[158]. When very little is known about the structure of an enzyme, *i.e.* only an amino acid sequence with no crystal structure or homology model, the screening effort can be vast, with every amino acid (except the ATG start codon) open to mutagenesis. For a hypothetical

300 amino acid protein; a single random amino acid change equates to $6x10^3$ possible enzyme sequences, two changes generates $1.2x10^5$ possible variants and three mutations gives rise to $2.4x10^6$ possible variants. As a rule, to attain 95% confidence that all possible variant sequences have been screened following a mutation protocol calibrated to introduce one mutation per gene copy, three times the number of possible mutated gene sequences must be screened. Thus, $1.8x10^4$ colonies would need to be screened to have 95% confidence of having fully mutated and screened every base in a 300 amino acid protein[159]. Accordingly, these numbers quickly become too large to reasonably screen using a typical high throughput HPLC/GC-MS assay. In some cases, supported by the ever growing level of bio-informatics available, it is possible to cut screening down to much more reasonable numbers by focusing on residues that are likely to be involved in substrate binding/catalysis[160]. A single amino acid position located within the active site can be substituted with all other 19 amino acids and screened with 95% amino acid mutation coverage on a single 96 well plate[158]. If a quick screening assay is available; for instance a two minute HPLC run, this screening can be accomplished in under four hours. The random mutagenic gene libraries needed for directed evolution can be created by a variety of differing *in vivo* and in *vitro* methods[161, 162]. A brief description of the main techniques used for this is discussed below.

3.2.1, In vivo mutagenesis:

3.2.1.1, *E. coli* mutator strains:

In vivo mutagenesis involves mutating a recombinant plasmid that harbours the gene of interest while it is being propagated by a cell. This is often accomplished in an *E. coli* mutator strain, such as XL1-Red (Stratagene)[163]. This strain is deficient in three major DNA repair mechanisms; *mutS* (deficient in error-prone mismatch repair), *mutD* (deficient in 3'- to 5'-exonuclease of DNA polymerase III) and *mutT* (unable to hydrolyse 8-oxo-dGTP). These allow for the XL1-Red's natural mutagenesis rate to increase ~5000 fold above WT *E. coli*[164]. A difficulty in working with this strain is that introduced mutations are not restricted solely to the gene of interest. Mutations will accumulate in both the plasmid backbone, and the host's genomic DNA. As a result, potential libraries must be constantly isolated and fresh XL1-Red re-transformed with this library. Any potential mutation which increases enzymatic activity must also be cloned into a fresh vector backbone to rule

out origin of replication (ori)/promoter driven increases in expression, and thus non-specific activity. Although, any mutations in the ori/promoter region that increases expression would likely be interesting and may be of some use in an expression plasmid. The XL1-Red strain provides a simple method to generate mutagenic libraries, but more sophisticated methodologies that require less time and labour are readily available.

3.2.1.2, *In vivo* chemical mutagenesis:

In vivo mutagenesis can also be carried out using chemical mutagens such as ethyl methanesulfonate (EMS), however, this process is not suited to selectively mutating a gene of interest; it is much more suited to wide spread mutagenesis of an entire genome[165]. To carry out specific mutagenesis of a gene, a much higher degree of mutagenic control is needed. This can be easily and quickly achieved with several *in vitro* techniques.

3.2.2, In vitro mutagenesis:

In vitro techniques form the work horse for most directed evolution projects. Their main advantage is that mutagenesis can be limited to the gene of interest. They are performed outside of a living cell and give rise to the phrase "test tube evolution", another term used for directed evolution. Most of these techniques are in the form of PCR based protocols, although it is possible to carry out *in vitro* mutagenesis using chemical mutagens.

3.2.2.1, Error prone PCR:

Error prone PCR (epPCR) is one of the most used, safest, and simplest methods for *in vitro* mutagenesis[166]. *Taq* based PCRs can be made error prone by including Mn^{2+} in the reaction buffer in place of Mg^{2+} or by using synthetic *dNTPS* to incorporate mutations (8-oxo-dGTP and 2'-deoxy-p-nucleoside-5'-triphosphate (dPTP))[167, 168]. These steps cause point mutations throughout the PCR product (gene) that can then be cloned into a non-mutated vector and evaluated for activity in an appropriate screening system. Error prone PCRs using *Taq* appear simple and cheap, but there are issues with the bias of mutations it provides[169, 170]. This skewed bias can be overcome by using a commercial epPCR kit, such as the genemorph II (Stratagene). Genemorph II uses two enzymes (the one *Taq* under mutagenic conditions, the other a proprietary error prone polymerase) with opposite biases to provide a balanced mutagenic spectrum. The overall mutagenic rate can be

controlled by altering the number of PCR cycles and the concentration of initial template. This gives different levels of amplification and as a result, varying levels of mutagenesis. PCR products generated with this combination of error prone polymerases should possess a high variety of mutations[169]. The random introduction of point mutations would at first glance appear to be an ideal method for generating mutations within proteins, however, this method fails to account for the complexity of the DNA genetic code[170]. When a single base is mutated, there is only around a 50% chance of this giving rise to a useful, chemically altered amino acid. The other possibilities of introducing a destabilising amino acid (glycine/proline), introducing a stop codon, or causing no amino acid mutation are just as likely. A major step forward would be made if methodologies were developed that enabled the quick and simple mutation of two or three base runs randomly throughout a gene. A complex approach for this has been developed (RID, random insertion and deletion) but it is technically challenging requiring multiple ligations[171].

The beauty of genetic code is that vast amounts of information can be stored using only four "letters of the alphabet" – A,G,T and C[172]. Nature has achieved this by introducing degeneracy by means of sixty-four codons (3 bp "words") that allow for twenty proteinogenic amino acids and three stop codons to be encoded solely by these four bases. However, this degeneracy makes the process of mutagenesis intrinsically difficult, as multiple, related DNA sequences encode the very same or highly related amino acids. For instance, the final base on many codons can be changed without fear of changing the encoded amino acid. Due to this effect, DNA mutations are not guaranteed to result in amino acid changes, *e.g.* the codons GGC and GGG both encode glycine in *E. coli*[173]. As a result, certain amino acids require higher levels of mutagenesis than others for specific changes, such as lysine (AAA, or AAG), which requires three mutations to become cysteine (TGT, or TGC). In other cases, only one mutation is required, such as valine (GTT, GTC, GTA, or GTG) to alanine (GCT, GCC, GCA, or GCG). If an error prone PCR is carried out with three random bp mutations per kb, there is only a 0.00000000074 % chance a specific K→C mutation compared to 0.3 % chance for a specific A→V. The genetic code appears to have evolved in such a way that when mutations within an amino acid sequence do occur, they are most likely to introduce changes wherein the chemistry is changed very little, for instance alanine to valine, or aspartate to glutamate[174]. The problem of limited mutagenesis due to codon degeneracy affects

any mutagenic technique which introduces single random point mutations into a gene.

3.2.2.2, *In vitro* recombination:

An excellent mutagenic technique based on biological mutagenesis is that of *in vitro* recombination, or "DNA shuffling"[175-177]. During cell meiosis, the chromosomal DNA can undergo crossover events which result in recombination within the genes contained on those chromosomes[178]. This process is essential to allow for sexually reproducing organisms to exchange genetic material without the accumulation of deleterious mutations. The nature of this process has been simulated in a test tube by using homologous genes which now act as the chromosomes to be recombined. The natural recombinases that carry out the chromosomal exchange process have been replaced by endonucleases and a DNA polymerase[175]. The *in vitro* process involves digesting homologous genes into 10 to 50 bp strands with a non-specific Deoxyribonuclease (DNase, EC 3.1.21.1) and then assembling these randomly back together with a DNA polymerase (**Figure 33**). How well this reconstruction occurs depends largely on how much conserved sequence identity there is between the fragments from each homologous gene[179].

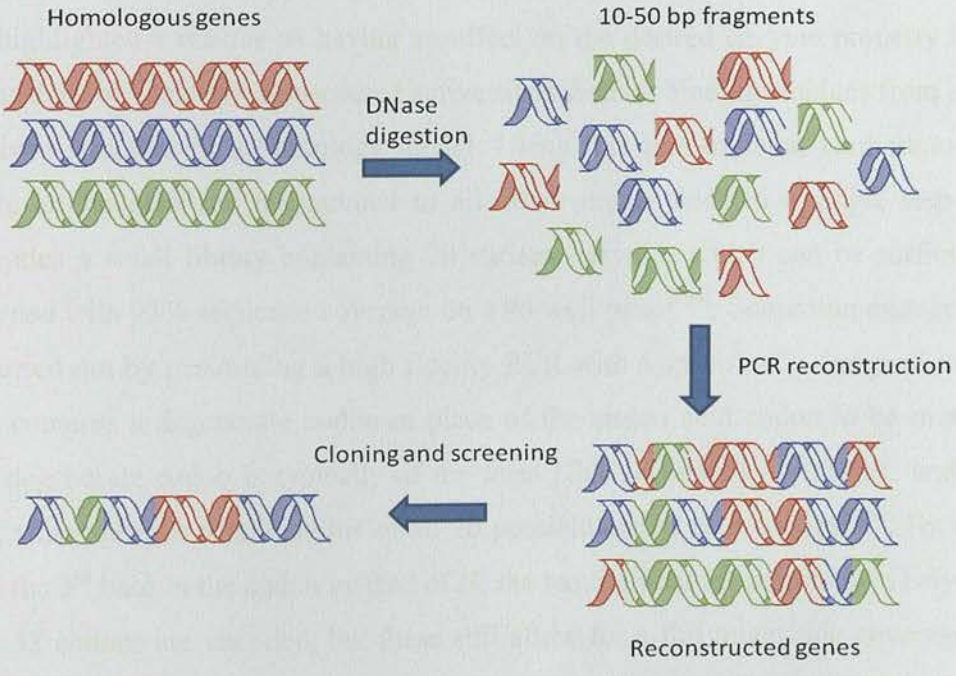


Figure 33. Digestion and random reconstruction methodology of DNA shuffling.

Using this method, point mutations should be kept to minimum level[180]. However, if desired, point mutations can be deliberately introduced into the process by using an

error prone polymerase, such as *Taq* with Mn^{2+} for the PCR reconstruction[181]. When point mutations are kept to a minimum, essential motifs within the sequence should be maintained *i.e.* conserved catalytic residues are not removed and structural motifs are in general left untouched. DNA shuffling was first reported in 1994 by Stemmer *et al*[175, 176]. Here, *in vitro* recombination was used to shuffle several TEM-1 β-lactamases and the reconstructed genes screened for improved resistance to cefotaxime. Three rounds of recombination resulted in a variant enzyme that increased resistance of an *E. coli* host by 32,000 fold. A similar effort using epPCR generated only a 16 fold improvement in resistance.

3.2.2.3, Saturation mutagenesis:

A third PCR based method used in directed evolution is that of saturation mutagenesis[182, 183]. This method is similar to epPCR, in that a PCR product is generated with point mutations, but very different in that the mutagenesis is controlled and selected for during the primer design process. Unlike epPCR, saturation mutagenesis does not suffer from limitations due to the degeneracy of the genetic code; instead it embraces it and uses only three point mutations to introduce 20 different possible amino acid mutations at a single position within the protein. Ideally, the process should be guided by initial epPCR/XL1-Red mutagenesis that has highlighted a residue as having an effect on the desired enzyme property being evolved, or by identifying suspected active site/substrate binding residues from either a solved structure or a homology model. Using this semi-rational methodology, a single amino acid can be mutated to all other amino acids in a single step; this generates a small library containing 20 variant enzymes which can be sufficiently screened with 95% sequence coverage on a 96 well plate[158]. Saturation mutagenesis is carried out by performing a high fidelity PCR with a specifically designed primer than contains a degenerate codon in place of the amino acid codon to be mutated. The degenerate codon is typically of the form NNK, where N = A/T/G/C and K = T/G, which allows combinations of all 20 possible amino acid codons[184]. By using K as the 3rd base in the codon instead of N, the requirement for screening is halved as only 32 codons are encoded, but these still allow for a full mutagenic coverage (all 20 amino acids). The PCR can be carried out in a site directed mutagenesis manner with complementary degenerate primers (with the reverse primer containing a complementary 5'-MNN-3' degenerate codon) or as a conventional PCR with a single degenerate primer and non-degenerate primer located elsewhere in the

gene/plasmid. In the latter case, a megaprimer is generated that can then be incorporated to the vector *via* restriction digest/ligation or MEGAWHOP PCR (**Figure 34**)[185-187].

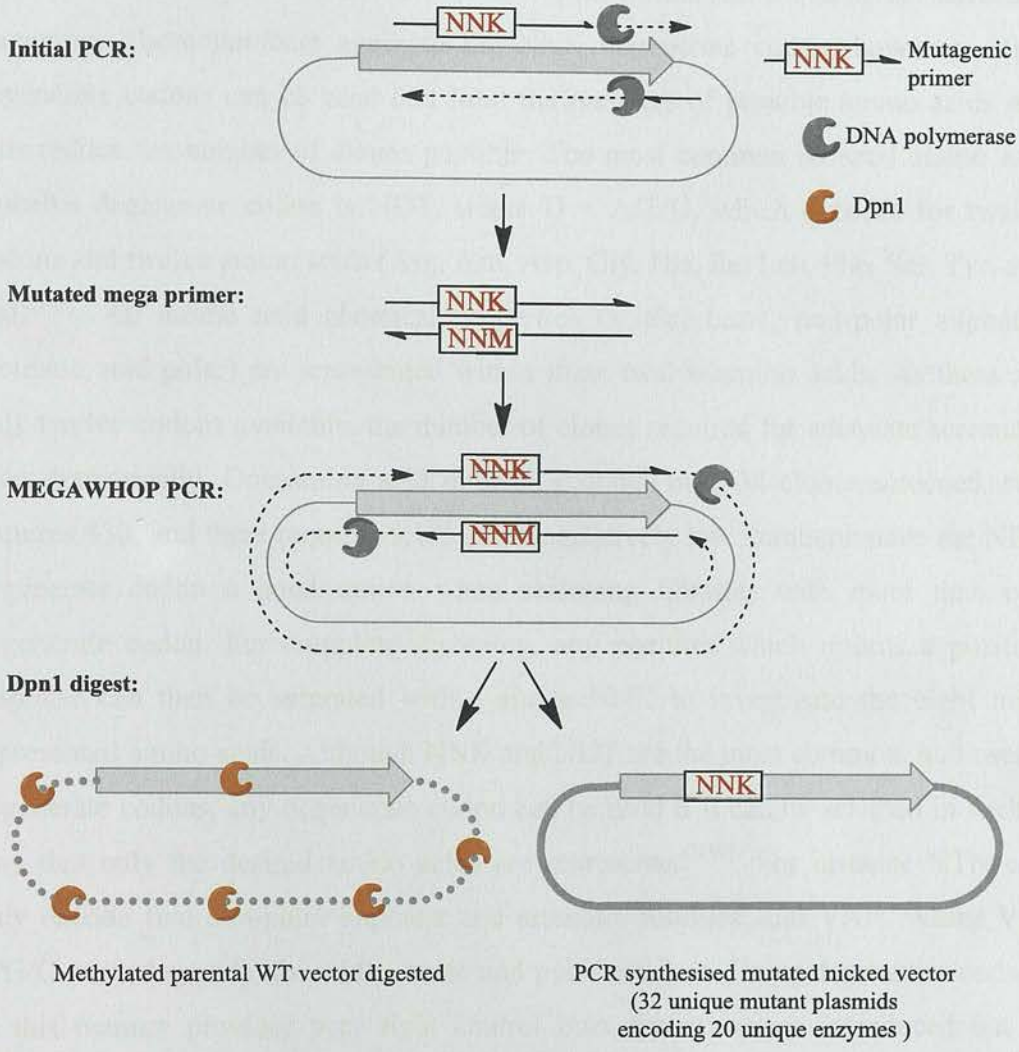


Figure 34. Saturation mutagenesis methodology using *DpnI* to digest the parental template.

Careful design of the primer(s) used can allow for a high degree of control over the mutagenesis process, for instance two amino acids can be simultaneously mutated by including two degenerate codons in a single primer. This approach is limited to amino acids that can be encoded on one primer[185]. Another approach which is of more use when the targeted residues are further apart is to use two degenerate primers to generate a megaprimer which spans two separate locations within the gene. In theory, another degenerate codon containing primer could then be used to extend this megaprimer allowing for three degenerate codons in a single PCR product. However, the chances of the final PCR product showing a full mutagenic spectrum at all three positions are low and the complexity of the generated library

becomes so great that effective screening becomes a major problem. When two positions are mutated using NNK degeneracy instead of only one, the number of clones that need to be screened for 95% coverage jumps from 94, to 3,066[158]. If this were increased to three codons, 98,163 clones would be required for effective screening. These numbers apply to the NNK degenerate codon, however, other degenerate codons can be used that limit the numbers of possible amino acids and thus reduce the number of clones possible. The most common reduced amino acid alphabet degenerate codon is NDT, where D = A/T/G, which encodes for twelve codons and twelve amino acids (Arg, Asn, Asp, Gly, His, Ile, Leu, Phe, Ser, Tyr, and Val)[188]. All amino acid chemical properties (acidic, basic, non-polar aliphatic, aromatic, and polar) are represented within these twelve amino acids. As there are only twelve codons available, the number of clones required for adequate screening drop dramatically. One amino acid mutated requires only 34 clones screened, two requires 430, and three requires 5,175. These relatively low numbers make the NDT degenerate codon a good option when screening libraries with more than one degenerate codon. For complete screening, any position which returns a positive response can then be saturated with a single NNK to investigate the eight non-represented amino acids. Although NNK and NDT are the most common, and useful degenerate codons, any degenerate codon can be used if it can be selected in such a way that only the desired amino acids are represented[189]. For instance NTN can only encode five non-polar aliphatic and aromatic residues, and VAN, where V = A/G/C, would specify six acidic, basic and polar residues. Using degenerate codons in this manner provides very tight control over the mutations introduced but is primarily of use when trying to limit screening to reasonable numbers with multiple degenerate codons, and is vulnerable to the degree by which desirable mutations (*i.e.* making a position of interest acidic will have the desired effect) can be predicted with a reasonable level of confidence.

3.2.2.4, Iterative Saturation mutagenesis:

Saturation mutagenesis is widely applied in an iterative form[160, 184, 188]. In this method, several sites (single or double amino acids) are mutated in separate libraries and the best performing variant from each library is re-mutated using the other targeted sites. In this way, four hypothetical sites - A/B/C/D, would all undergo saturation mutagenesis to give four separate libraries[190]. If the best performing variant was found in site A, this would then be saturated at positions B/C/D to give

three further libraries. Again the best variant (*e.g.* from library B containing the previous A library mutation), is used for mutagenesis at C/D giving two libraries. The process is repeated until an optimised variant has been found for all four sites (**Figure 35**). Screening in this way allows for four sites to be mutated with screening of ten libraries each containing only 100 colonies with 95% mutagenic coverage. Screening these sites in one library would not be realistically feasible – if mutagenesis could somehow be limited to the twelve bases of interest, $>3x10^6$ clones would need to be assayed. At any stage in the A→B→C→D process, any position can then be re-screened with relative ease to include further library diversity in the screen, and would only require screening of an extra 100 colonies per site. This method drastically cuts down on the total number of colonies needed to be screened and has been termed combinatorial active site saturation testing (CASTing)[160, 191]. Of course, to saturate the active site efficiently, the amino acids that make it up must be known, which often requires a solved crystal structure of the protein of interest or a high homology model, which is not always immediately available.

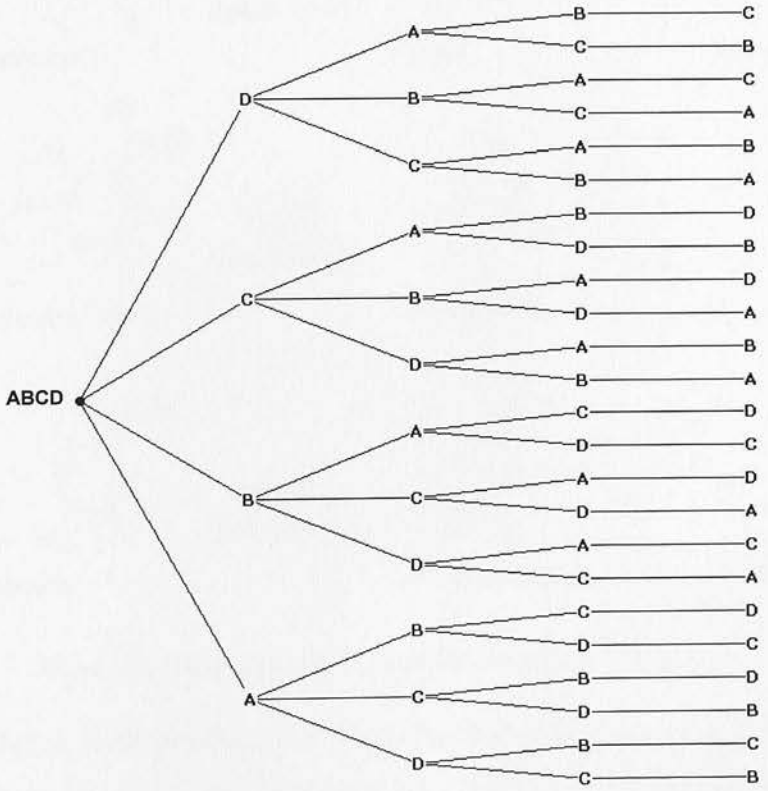


Figure 35. Iterative screening of four saturation positions (A, B, C, and D).

3.2.2.5, Chemical mutagenesis:

A less common method for *in vitro* mutation is that of chemical mutagenesis[192]. However this approach has been used very successfully in the past for generating libraries with a high variety of mutations[193]. Several compounds are available which can be applied directly to DNA that will induce point mutations (mutagens). Examples of these mutagens are ethyl methanesulfonate (EMS), nitrous acid (HNO_2), and hydroxylamine (NH_2OH) (**Figure 36**)[161, 192, 194] 138].

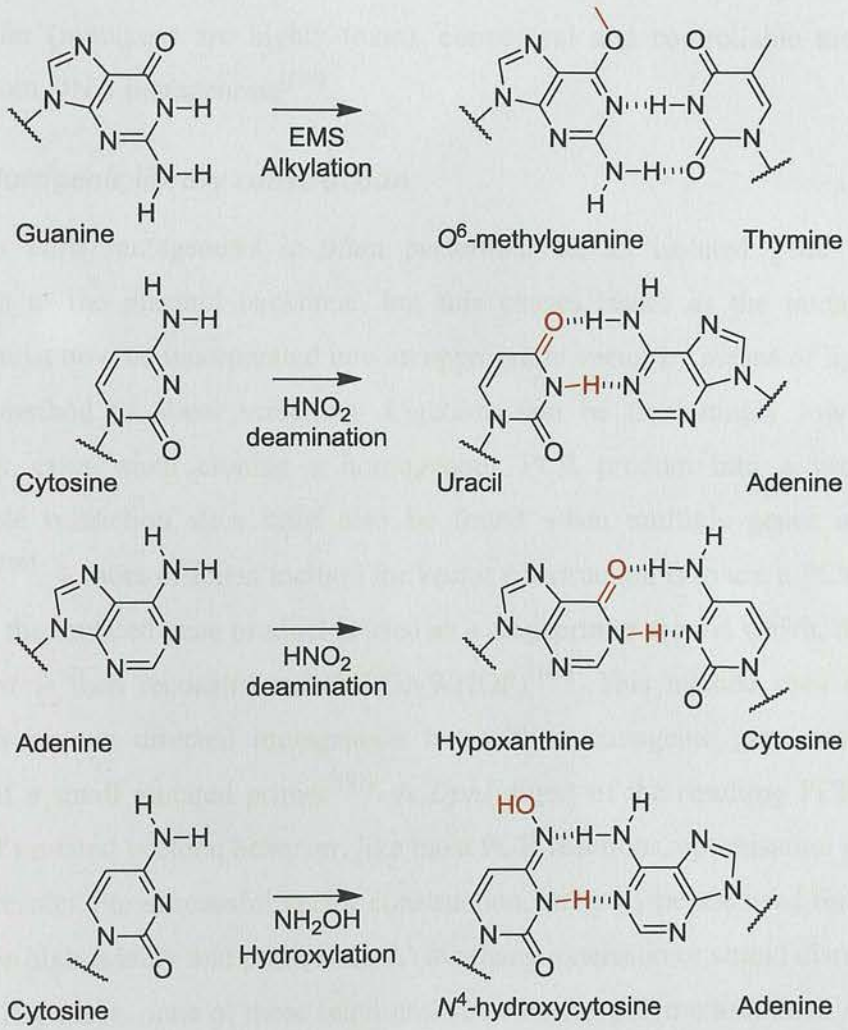


Figure 36. Mis-match effects of chemical mutagen mutagenesis.

EMS introduces point mutations by alkylating the keto group of guanine so that it can now mismatch base pair with thymine. This alkylation can cause G/C→A/T transitions (substitution of purine → purine or pyrimidine → pyrimidine) during DNA replication. Other alkylating agents such as methyl methanesulfonate or dimethylsulfonate can also be used, and in general they have similar mutagenic properties to EMS. HNO_2 carries out oxidative deamination of both adenine and

cytosine to introduce G/C→T/A transversions (substitution of purine → pyrimidine or vice versa) during DNA replication. The deamination of adenine and cytosine generates hypoxanthine and uracil respectively. Guanine also undergoes deamination to xanthine but as this can still pair with cytosine as the usual Watson-Crick base pairing still occurs. NH_2OH reacts preferentially with cytosine to give N^4-hydroxycytosine which can mismatch with adenine to give G/C→T/A transitions. Nowadays these compounds are more likely to be used by those carrying out mutagenesis in model organisms than isolated DNA as PCR techniques provide a much safer (mutagens are highly toxic), convenient and controllable method for carrying out DNA mutagenesis[195].

3.2.3, Mutagenic library construction:

In vitro mutagenesis is often performed on an isolated gene to avoid mutations to the plasmid backbone, but this causes issues as the mutated PCR product must now be incorporated into an appropriate vector by means of ligation, or another method to allow screening. Ligations can be frustratingly low in their efficiency even when cloning a homogenous PCR product into a vector, and compatible restriction sites must also be found when multiple genes are being shuffled[196]. A more efficient method for vector construction is to use a PCR process in which the mutated gene product is used as a megaprimer around which, the rest of the vector is then reconstructed (MEGAWHOP)[186]. This method uses the same principles as site directed mutagenesis but with a mutagenic gene megaprimer instead of a small mutated primer[197]. A *DpnI* digest of the resulting PCR gives a library of mutated vectors; however, like most PCR reactions, optimisation is needed for best results. For successful vector construction, any polymerase used for this step must have high fidelity and possess no A' overhang extension or strand displacement activity. *Taq* meets none of these requirements; whereas polymerases such *Pfu turbo* and *pfx platinum* match them all, making them ideal polymerases for most mutagenic library work ups.

3.3, Selecting and screening variant libraries:

To increase the likelihood of directed evolution being successful, sampling high volumes of sequence space with ease must be possible. This can be done in two ways, either by *in vivo* selection/screening methods, or *in vitro* high throughput screening. *In vivo* selection can be used to evaluate library sizes in the millions using

genetically modified bacterial cell lines which rely on the evolved reaction to provide an essential compound for life or remove a toxic one. Using *in vivo* screening, libraries of $>10^4$ c.f.u. can be screened on an agar plate using a visual indicator of enzyme turnover. Using *in vitro* screening, multiples of 96 well plates can be screened with almost analytical accuracy

3.3.1, In vivo selection/screens techniques:

3.3.1.1, *In vivo* selection:

In vivo assays can be in the form of either a genetic selection or high throughput screening – the difference being that in a selection method, poorly performing variants should be removed from the background by host death, whereas in a screening approach, even catalytically dead variants will be assayed. *In vivo* selection methods can be used to screen vast libraries and typically give a life/death or growth/no growth assay in which an increase in the desired characteristic helps the host bacteria to survive and/or possibly out-grow its competitors[198]. *In vivo* selections can suffer from a lack of quantification that means potential hits often need to be re-screened using more accurate analytical methods such as HPLC/GC-MS. They can be used as an initial screen to remove redundant sequences with no activity and subsequently cut down on the variants that need screened *in vitro*[199]. The major advantage of *in vivo* selection is that the size of libraries that can be screened is in the millions[200]. An classic example of *in vivo* selection is evolving resistance to an antibiotic – for instance Stemmers evolution of cefotaxime resistance using DNA shuffling (See section 3.2.2.2)[175]. *In vivo* selection in auxotrophic bacteria can be a very powerful process if a suitable auxotroph can be engineered that is complemented by product of the reaction of interest. Kagamiyama *et al* have utilised an *E. coli* strain deficient in a Branched-chain amino acid aminotransferase (BcaT, EC 2.6.1.42) required for the final stage of L-valine, L-leucine, and L-isoleucine biosynthesis to increase the branched-chain substrate acceptance of Aspartate aminotransferase (AspTA, EC 2.6.1.1). The strain was transformed with an AspTA mutagenic library and grown on minimal media supplemented with 2-oxovaline, L-leucine, and L-isoleucine. Repeated selection of the fastest growing colonies, *i.e.* those that could transaminate 2-oxovaline effectively, revealed a variant AspTA with 10^5 fold increase in activity towards branched-chain amino acids compared to the WT[201]. *In vivo* selections can be difficult to devise as the internal

chemical pathways of bacteria are complicated and inter-connected, with many "escape routes" to any applied selection being possible.

3.3.1.2, *In vivo* screening:

In vivo screening utilises a visual assay to indicate improvements in enzyme activity[202]. This may be relatively straightforward if the product can be directly detected (*e.g.* coloured products) or can be fed easily into a coupled enzyme assay which gives a coloured response. For instance, the synthesis of a D-amino acid can be detected in a bacterial colony on an agar plate using a combination of D-amino acid oxidase (D-AAO, EC 1.4.33) and Horseradish peroxidase (HRP, EC 1.11.1.7). The D-amino acid is oxidised to the keto acid by D-AAO releasing H_2O_2 which HRP uses to oxidise a redox indicator added to the agar giving a coloured colony[203]. As D-amino acids are not found in *E. coli* at high concentrations, any colony possessing the ability to produce these (from an appropriate substrate) will be coloured more so than its neighbours and should be easily identified.

3.3.2, In vitro screening techniques:

In vitro screening ideally takes place in 96 or even 384 well plates and uses an optical assay (UV/visible/fluorescence) to measure enzyme activity[204]. This process begins with small scale growth of colonies containing variants in 96/384 well plates followed by enzyme expression, cell lysis and clarification. The cell free extract (CFE) can be used as source of enzyme activity to test for improved variants. This has the added advantage that it screens for soluble/folded protein as well as activity[205]. Detection of enzyme turnover can be carried out by any number of chemical assays, and this step depends entirely of the nature of the substrate/product and any organic cofactor used. Examples of possible assays included using D/L-AAO/HRP, or by measuring the rate of NADPH/NADP oxidation by a decrease in absorbance at 340 nm[206-208]. Often, coupled enzyme systems are used to cause a measurable change, such as the use of Pyruvate kinase (PK, EC 2.7.1.40) and Lactate dehydrogenase (LDH, EC 1.1.1.27) to measure ATP hydrolysis. Here, the hydrolysis of ATP triggers the reduction of NAD^+ which can be monitored by an absorbance increase at 340 nm[209]. In these assays, care must be taken to ensure that the rate limiting step is the reaction being investigated to allow for increases in this step to be measured accurately. Once an effective *in vitro* screen has been developed, multiple 96/384 well plates can be screened rapidly with the use of a liquid handling robot and

a plate reader. If a coupled UV/vis/fluorescence assay cannot be adequately designed for a given reaction, a high throughput HPLC/GC-MS/LC-MS assay can be used instead. However, these are less than ideal for high-throughput screening as even with a quick two minute sample/detection time, a 96 well plate will still require three hours for full analysis, and given the requirement for high-purity samples, extra liquid handling steps may be required beforehand. One clear advantage of *in vitro* screening over *in vivo* techniques is that they can be carried out in semi-quantitative manner, whereas *in vivo* methods are more qualitative but higher throughput.

3.3.3, Fluorescence assisted cell sorting:

Fluorescence assisted cell sorting (FACS) matches the sensitivity of *in vitro* screening with the throughput of *in vivo* selection is Fluorescence-activated cell sorting (FACS)[210, 211]. This method allows for high cell densities (>10^7 cells) to the screened (sorted) depending on the fluorescence of individual cells. If an assay can be developed in which the product is fluorescent, the reaction can be directly assayed and hits physically separated from the non-hits. FACS can also be applied as a life/death assay in which a living cell activates a fluorescent dye, whereas a dead cell does not. In this situation the reaction being evolved must generate either a required metabolite due to a desired reaction (*e.g.* novel activity on the desired non-natural substrate) or a toxin if the reaction is undesired (*e.g.* activity on the non-desired natural substrate)[212]. Ideally, both these selective pressures are applied at the same time otherwise non-active enzymes would be selected as false positives due lack of toxin release. FACS appears to be an attractive way to perform directed evolution, but it is hindered by the requirement for a fluorescent product/life death assay and access to an expensive and mechanically complicated fluorescence-activated cell sorter.

3.4, Computer aided directed evolution:

Computational *de novo* design of novel enzymes to carry out required reactions efficiently would be the pinnacle of protein engineering[213]. Although whether this constitutes directed evolution is debatable, the end goal is the same, with a computer now conducting *in silico* screening of hypothetical mutations and protein folds to find an ideal arrangement of amino acids. The use of *in silico* tools to identify biocatalysts by sequence analysis has already been shown, the next step would be the *de novo* design of such enzymes[214]. The timeframe in which this will

become routine (if at all) is difficult to judge as our understanding of how protein sequence dictates folding and activity is poor[215, 216]. Dywer *et al* infamously claimed to have designed a *de novo* enzyme capable of triose phosphate isomerase activity. However, it was later discovered that a low level *E. coli* Triose phosphate isomerase (TIM, EC 5.3.1.1) contaminant was in actual fact responsible for the observed activity[217, 218]. The contaminating TIM was believed to have been co-purified during the nickel resin step, at a concentration which gave the appearance of low level activity from a novel enzyme. This is unsurprising given the high catalytic efficiency of TIM and highlights the need for effectively "nullified" background reactions in purification hosts when performing this type of experiment[219]. The recently reported computational design of enzymes capable of catalysing a Kemp elimination (Kemp eliminase) and a Diels-Alder reaction (Diels-Alderase) has shown that this field of protein engineering can develop enzymes not only carrying out the required chemistry, also entirely novel enzymatic reactions **(Figure 37A** and **Figure 37B)** [220, 221]. Whether the Diels-Alder reaction is used by nature or is truly novel is still an unsolved question, although evidence would suggest that fungal Lovastatin nonaketide synthase (LNS, EC 2.3.1.161) may perform an intra-molecular Diels-Alder reaction[222, 223]. Certainly, the engineered intermolecular Diels-Alderase would appear to be novel. Some of the earliest directed evolution work using catalytic antibodies did succeed in developing proteins that could carry out the Deils-Alderase reaction but they were exceptionally poor and this approach is no longer in use[224]. At present, *in silico* designed enzymes invariably suffer from hindered activity that is several orders of magnitude lower than those of their naturally evolved counter-parts, but by using directed evolution, improvements can be made in the designed protein. The catalytic efficiency of a Kemp eliminase was improved over 200 fold by directed evolution after an initial *in silico* design. The possibility of bespoke enzymes carrying out custom reactions may not ever be truly realised due to the complicated nature of enzyme design, but so far, impressive steps have been made towards this goal.

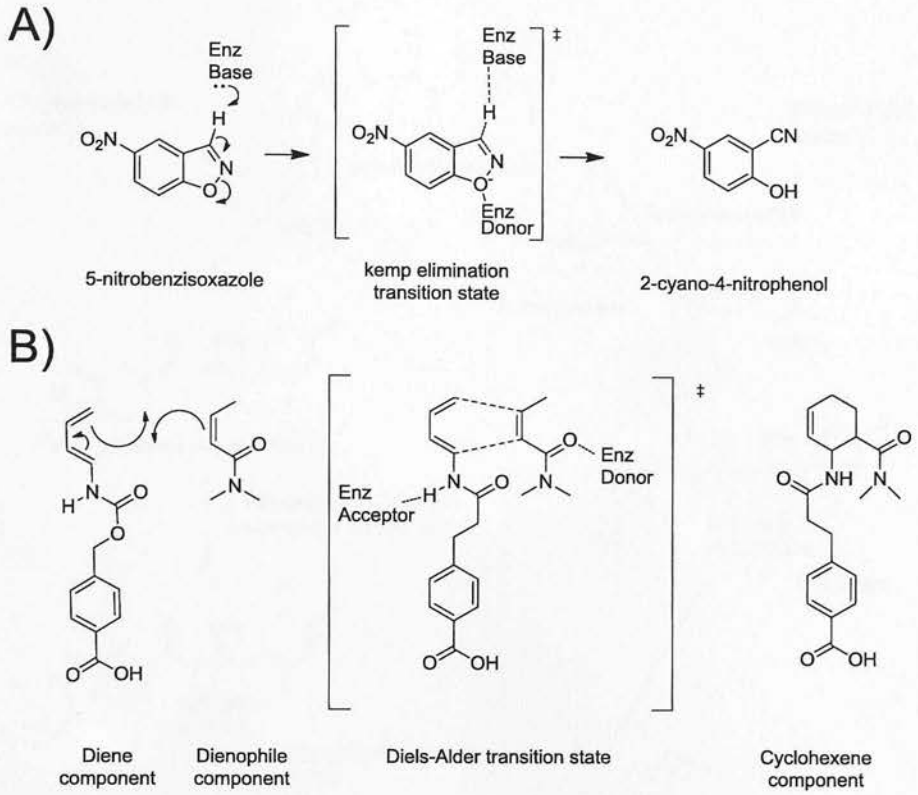


Figure 37. Novel enzymatic chemistry designed *in silico*. A) Kemp elimination catalysed by an enzymatic base, and stabilised by an enzyme H bond donor. B) Intermolecular Diels-Alder activity exhibited by designed enzyme with H-bond donor and acceptor stabilisation.

3.5, Directed evolution of a Sitagliptin transaminase:

The directed evolution of a transaminase biocatalyst for the production of sitagliptin, developed by Codexis and Merck, is an extremely elegant example of evolution being used to overcome the inherent failings of a possible industrial biocatalyst[18, 225]. The biocatalytic synthesis replaced an asymmetric hydrogenation using a Rhodium catalyst at 250 psi. Along with a cleaner and cheaper route to the desired product, the % e.e. was increased from 97% to >99.9%, and the yield increased from 82% to 91% (**Figure 38**).

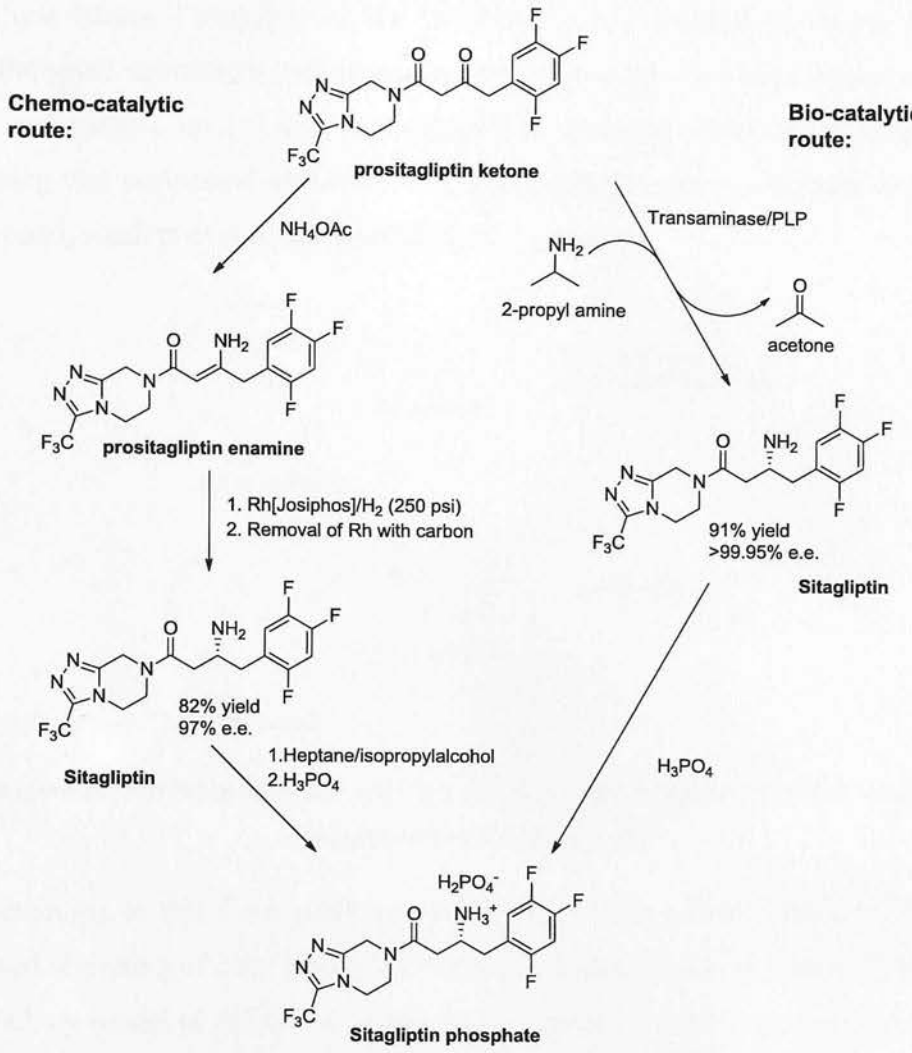


Figure 38. The chemo- and bio- catalytic routes to sitagliptin.

This work is marked for several reasons, such as the evolution of not just activity on a novel substrate but also constant engineering of the protein stability towards the final required process conditions. The WT enzyme was stable under the mild conditions of 5% DMSO at 22°C, pH 7.5 in the presence of 0.5 M 2-propyl amine (the amine donor in the reaction) whereas after eleven rounds of evolution, it was stable at 50% DMSO at 45°C, pH 8.5 in the presence of 1 M 2-propyl amine. Each of these parameters was increased gradually with each round of screening to allow for the final protein to be industrially useful. Directed evolution was performed on a PLP dependent transaminase (ATA-117) from *Arthrobacter sp.* that had previously been shown to carry out (*R*)-specific reductive aminations of small methyl and cyclic ketones[226]. A major problem that had to be overcome was the traditional large pocket/small pocket substrate specificity of transaminases[227, 228]. The two R group pockets within the enzyme active site limit the substrate range to those that can fit

into these spaces. Prositagliptin, the substrate for this biocatalyst, clearly falls out with the small substituent/pocket requirement (**Figure 39**). The large pocket was first optimised using a methyl ketone prositagliptin analogue. After this was achieved, screening was performed with tri-fluoro benzyl substituted prositagliptin to open up the second, small pocket to accommodate this group.

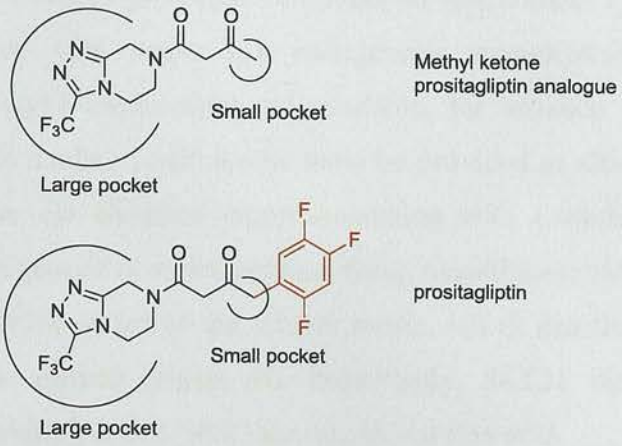


Figure 39. Screening approach used to bypass the large pocket/small pocket binding characteristics of transaminases.

The screening in this work made use of HPLC and LC-MS *in vitro* screening and involved screening of only 10,000 clones due to semi-rational mutagenesis based on a homology model of ATA-117. Attempting to optimise both pockets at once would have required a substantially higher level of screening. The work shows that effective industrial biocatalysts can be engineered that not only rival, but out perform their equivalent chemo-catalytic routes.

3.6, Directed evolution of NAAAR:

To allow directed evolution of NAAAR, bacteria will be grown in conditions where selective pressure will favour those that can racemise *N*-acetyl amino acids. To apply this selective pressure, a double knockout bacterial host (SET21) will be used allowing for *in vivo* selection of large random libraries generated with epCPR[229]. High-throughput screening using Chiral HPLC will also be used to screen focused saturation mutagenesis libraries of mutation hotspots. The SET21 strain was developed by Dr Sylvain Royer and highlighted a NAAAR enzyme containing a G291E mutation that is believed to improve NAAAR activity. The SET21 strain is deficient in the *de novo* L-methionine pathway (from L-homoserine), *via* a Cystathionine-γ-synthase knockout (MetB, EC 2.5.1.48), and a D-methionine racemisation pathway by means of a D-amino acid dehydrogenase knockout (DadA,

EC 1.4.99.1)[230-234]. These knockouts were achieved using the published PCR product gene knockout method of Datsenko *et al*[235]. The gene encoding MetB was replaced with a chloramphenicol resistance marker and the gene encoding DadA was deleted. Deletions removed any chance of reversion to the WT genotype, and insertion of an antibiotic resistance marker allowed for strain selection. These genetic modifications limit SET21 growth to conditions in which either L-methionine or *N*-acetyl-L-methionine (the strain has endogenous aminohydrolases capable of hydrolysing *N*-acetyl-L-methionine) are available, for instance LB media. When grown on minimal media, L-methionine must be provided to allow SET21 growth. This can be done *via* chemical complementation with L-methionine/*N*-acetyl-L-methionine, or *via* genetic complementation using recombinant NAAAR to racemise *N*-acetyl-D-methionine added to the growth media, which can then be used by the strain to facilitate growth (**Figure 40**). Importantly, SET21 cannot grow on D-methionine and neither can NAAAR racemise this amino acid.

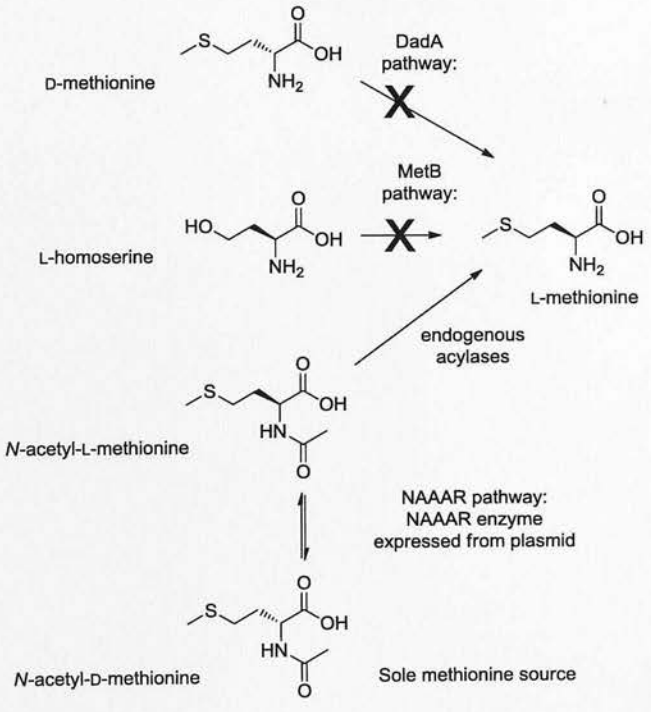


Figure 40. Disabled L-methionine biosynthesis and genetic NAAAR complementation.

By creating a system is which NAAAR turnover is the rate limiting step for strain growth, NAAAR variants will be effectively assayed *in vivo* and those with the highest activity should outgrow their neighbours. This dependence on NAAAR activity will allow for directed evolution to be applied, and the resulting variants

which will hopefully contain beneficial mutations can be selected and tested as possible biocatalysts.

Chapter 4, Directed evolution of Amycolatopsis Ts-1-60 NAAAR:

"I'm simply saying that life, uh..... finds a way"

Dr Ian Malcolm (Jurrasic Park)

4.1, Directed evolution of Amycolatopsis Ts-1-60 NAAAR:

Directed evolution allows for new functions to be introduced, or the improvement of existing ones, in an enzyme/protein[236]. Using this technique, the promiscuous *N*-acetyl amino acid racemase activity of NAAAR from *Am. Ts-1-60* has been enhanced. This process involved using the SET21 host to screen variant NAAAR plasmid libraries for mutations that caused desirable characteristics.

4.2, SET21 microbiology:

4.2.1, SET21 host growth conditions:

Before screening of variant libraries could be performed, it was necessary to become familiar with the SET21 host and the microbiology techniques required for its use. The growth requirements for the host (with, and without NAAAR present) were confirmed with growth of SET21(DE3) expressing either empty pET20b or pET20b NAAAR WT on minimal media containing either L-methionine, D-methionine, *N*-acetyl-L-methionine, or *N*-acetyl-D-methionine (all 2.5 mM). As expected, and shown by Dr Royer, the SET21 cell line requires a source of L-methionine for growth (**Figure 41**). SET21(DE3) transformed with pET20b did not grow when streaked on D-methionine or *N*-acetyl-D-methionine, whereas growth was confluent on media containing L-methionine and *N*-acetyl-L-methionine. Growth on *N*-acetyl-L-methionine confirms that this host possess an L-acylase capable of hydrolysis of *N*-acetyl-L-methionine to the release the free amino acid. SET21(DE3) expressing NAAAR from pET20b NAAAR WT again showed no growth on media containing D-methionine and confluent growth on *N*-acetyl-L-methionine/L-methionine. Importantly, these cells were now capable of growth on *N*-acetyl-D-methionine due to NAAAR racemising this to the L-enantiomer. No growth was observed with either host on D-methionine confirming the requirement for D-methionine to be *N*-acetylated for enzymatic racemisation.

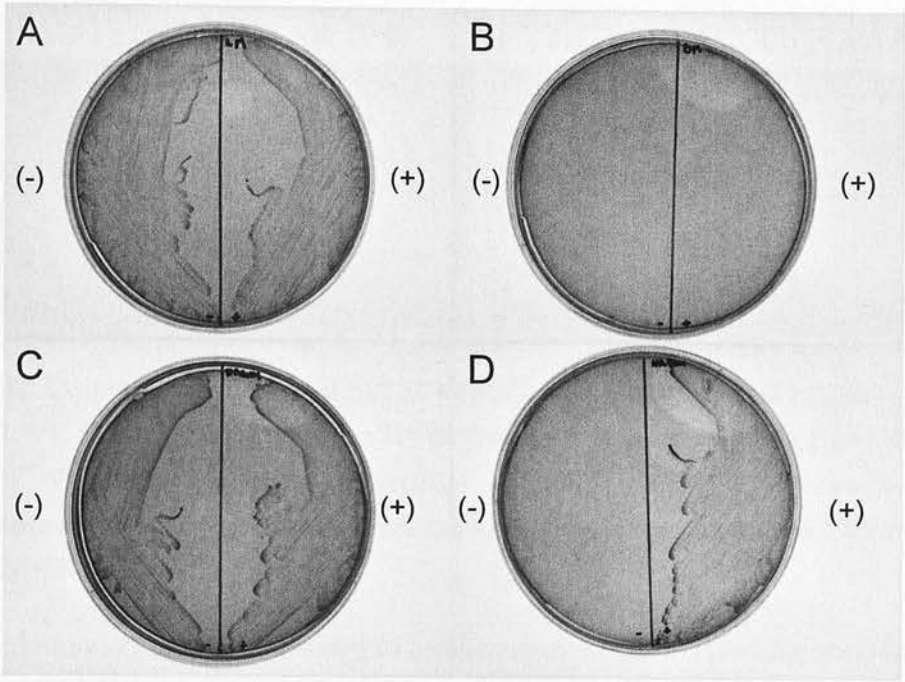


Figure 41. Growth conditions of SET21(DE3) with pET20b = (-) and pET20b NAAAR = (+). A = L-methionine, B = D-methionine, C = *N*-acetyl-L-methionine, D = *N*-acetyl-D-methionine. Bacterial growth is stained with 2,3,5-Triphenyltetrazolium chloride redox indicator giving bactreia a dark contrast on the light agar plate.

4.2.2, Improved expression system for screening:

For saturation mutagenesis of G293, NAAAR was expressed in the SET21 (non DE3) host using a pTTQ18 plasmid, driven by a *tac* promoter[237]. This strong promoter is a hybrid of the separate *lac* and *trp* promoters used for recombinant protein expression. The promoter is followed downstream by the *lac* operator and as the plasmid also contains the $lacI^Q$ gene, protein expression is very tightly regulated. The addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) is required to overcome the $lacI^Q$ repression and allow protein expression. This system had proven to be effective in the past with discovery that a G293E mutation in NAAAR increases activity, and was used with success in saturation mutagenesis of G293 (See section 3.2.2). However, it gave non-reproducible growth during early stages of testing. Difficulties were encountered as when the SET21 host was transformed with pTTQ18 NAAAR WT and spread on M9 minimal media containing *N*-acetyl-D-methionine (2.5 mM) and IPTG (1 mM), the resulting colonies would show a considerable variation in size. In comparison, SET21(DE3) transformed with pET20b NAAAR WT and spread on Davis minimal agar (DMA) containing 2.5 mM *N*-acetyl-D-methionine with no IPTG showed a more uniform colony size (**Figure 42**).

SET21 pTTQ18 NAAAR WT: SET21(DE3) pET20b NAAAR WT:

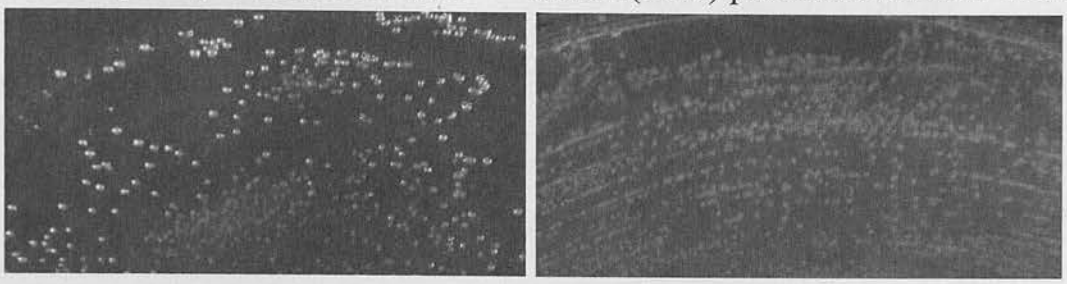


Figure 42. Growth characteristics of SET21 pTTQ18 NAAAR WT and SET21(DE3) pET20b NAAAR WT. SET21 pTTQ18 grown on M9 minimal agar supplemented with 1 mM IPTG and 2.5 mM *N*-acetyl-D-methionine. SET21(DE3) pET20b NAAAR WT grown on DMA supplemented with 2.5 mM *N*-acetyl-D-methionine. Uniform colony size with pET20b compared to variable colony size with pTTQ18.

This variation in size was attributed to the requirement for IPTG to induce expression of NAAAR in the pTTQ18 plasmid to facilitate growth. The concentration of IPTG in any given cell is related to the number of active lactose permeases available for uptake of IPTG and is controlled by stochastic fluctuations in protein expression[238, 239]. Higher numbers of lactose permeases gives higher levels of expression, and *vice versa*. Further variation in the exact cellular inducer levels can arise due to Lactose permease expression/activity being passed on to daughter cells by asymmetric cellular division and other pre-induction genetic/phenotypic dispositions. Given that the rate of SET21 growth is dependent on NAAAR activity; any cells which happen to have higher levels of NAAAR expression than background will outgrow their neighbours. Further to this, cells which undergo IPTG induced expression do so under enormous metabolic burden compared to those that do not[240]. A combination of effects may lead to the differing colony sizes observed with SET21 expressing a single NAAAR enzyme from the pTTQ18 plasmid on minimal media containing IPTG (**Figure 42**). It should be noted that although cell size of any given SET21/NAAAR population varied, in general, SET21 cells expressing NAAAR G293E were larger than those expressing NAAAR WT. To remove the variation in colony size arising from IPTG effects, the gene encoding the NAAAR protein was cloned into the multiple cloning site of pET20b *via* the NdeI and HindIII sites. The cloning of this IPTG independent, NAAAR expression plasmid was performed by amplification of NAAAR WT gene from pTTQ18 NAAAR WT using primers *NAAAR NdeI F* that encoded an *N*-terminal NdeI site and *NAAAR HindIII R* that encoded a *C*-terminal HindIII site. As the native start codon for *Am. Ts-1-60*

NAAAR (GTG) is retained in the pTTQ18 plasmid, this was mutated to ATG to allow better expression in *E. coli*. The restriction sites are shown below in bold, with the mutation underlined:

NAAAR NdeI F 5'-**CAT <u>A</u>TG** AAA CTC AGC GGT GTG GAA CTG-3'
NAAAR HindIII R 5'-**AAG CTT** CTA CGA ACC GAT CCA-3'

The PCR product containing the NAAAR gene was gel extracted and ligated into the pGEM T-easy cloning vector. From here, the plasmid was digested with *NdeI* and *HindIII* before ligation into pET20b digested with the same enzymes. The sequence of the pET20b NAAAR WT clone (4715 bp) was confirmed by sequencing (gene only). This was repeated with the NAAAR G293E gene to generate pET20b NAAAR G291E. In going from pTTQ18 to pET20b, an EcoRI restriction site from a previous cloning step was removed from the 5' end of the sequence between bases 4 and 10. This step removed two non-native amino acids (Glu-Phe) from the translated protein sequence and accounts for the discrepancy in numbering between G293 in pTTQ18 and G291 in pET20b. This plasmid allows expression of native NAAAR from the strong T7 promoter but does not require the addition of IPTG to do so. To allow the use of the T7 promoter, the SET21 host was made DE3 lysogenic using a commercial kit and the manufacturer's instructions (Merck). The constitutively ("leaky") expression from pET20b occurs because there is no *lac* operator downstream of the T7 promoter to repress expression. IPTG is not required for expression of the T7 RNA polymerase from the SET21 chromosome (derived from BW23115) as the host is $lacI^{+}$[241]. The switch from pTTQ18 to pET20b gave a marked decrease in the deviation of colony size between SET21(DE3) colonies, with almost all colonies appearing to be the same diameter (**Figure 41**).

4.2.3, Minimal media:

The media used for SET selection was also modified from the original M9 minimal media previously used by Dr Royer. This media required M9 salts, $MgSO_4$, Glucose, $CaCl_2$, and $CoCl_2$ solutions to be separately prepared and sterilised[242]. These solutions were then quickly added to Agar Number 1, before being aliquoted into sterile Petri dishes. The process was open to contamination due to the number of components that had to be added. During early SET21 testing a large number of contaminations (bacterial growth on D-methionine/*N*-acetyl-D-methionine/no methionine) were encountered. In an attempt to remove these, the minimal media

being used was changed from M9 to Davis minimal media[243]. Similar to M9, Davis minimal media is a potassium phosphate buffered solution (pH 7.0) containing various compounds (magnesium sulphate, ammonium sulphate, glucose and sodium citrate) that can be assimilated by bacteria into all the components required for life. The media is available as a commercial broth or agar making its preparation a simple task with less chance of contamination or formulation mistakes. Contaminations did decrease as SET21 testing progressed but whether this was due to the change in media or improved microbiological techniques is unclear.

4.2.4, SET21(DE3) competency:

To perform effective directed evolution using the SET21 selection host, it was important to ensure that large numbers of plasmids could be screened on a relatively small number of agar plates. This required SET21 cells that were highly competent allowing for maximal screening per transformation. Initial attempts at making competent SET21 cells involved using the standard $CaCl_2$ technique[244]. This involved back diluting an SET21 overnight culture to an OD_{600} of 0.1 in 100 mL of LB broth (30 μg mL^{-1} chloramphenicol) and allowing this to grow with agitation at 37°C until the OD_{600} reached 0.5. After an hour on ice, the cells were harvested by centrifugation and treated with 100 mM $CaCl_2$. The final volume (~2 mL) was stored as 200 μL aliquots. A single aliquot would be frozen and then defrosted to check cell competency. This involved transforming the SET21 aliquot *via* the standard heat shock method with 0.01 ng of pUC18 (Invitrogen) and spreading serial dilutions of this on LB agar (100 μg mL^{-1} ampicillin). Five μL was then streaked on separate DMA plates containing *N*-acetyl-D-methionine, or L-methionine to check for contaminations. No antibiotics were added to these plates. Using this technique, cell competencies of $1x10^3$ – $1x10^5$ c.f.u. per μg pUC18 could be regularly attained. These values were deemed far too low for effective screening. As electroporation is known to generate much higher cell competencies than the $CaCl_2$ heat shock method previously used, focus switched to this technique[245]. The SET21 host was made electro-competent by culturing SET21 overnight at 37°C in LB broth lacking NaCl (30 μg mL^{-1} chloramphenicol) and sub-culturing back to an OD_{600} of 0.1 in to 100 mL of the same media. The procedure was based on that reported by *Shi et al*[246]. Once these cells had reached an OD_{600} of 0.5, they were harvested by centrifugation and washed twice with 25 mL ice cold ddH_2O (followed by 10 mL of 20% glycerol), and finally centrifuged and re-suspended (gentle

swirling for all resuspension steps) in 1 mL of ice cold 20% glycerol, 150 mM trehalose. The final OD_{600} of this 1 mL sample was ~100. This was stored as 50 µL aliquots, and was tested for both competency with 0.01 ng pUC18 (Invitrogen) and contamination as above. Using this technique, cell competencies between $1x10^8$ - $1x10^9$ c.f.u. per µg pUC18 could be regularly attained, with occasional values $>1x10^{10}$ c.f.u. per µg pUC18 being measured. The inclusion of NaCl in the growth media was found to drop competency to ~$1x10^7$ c.f.u. per µg pUC18. Keeping the cells below 4°C was believed to be of key importance. Although no studies were performed to address the importance of trehalose in the final re-suspension stage, literature evidence suggests it acts as a biological stabiliser of both protein and lipids and is most likely an important component for obtaining high competencies. It is often included in bacterial freeze drying formulations to help protect the integrity of cellular membranes, and is possibility performing a similar role here given the high stresses exerted on the cell during the electroporation process[247, 248].

4.3, Mutagenesis:

Throughout this project, various means of mutagenesis have been used for both rational and random mutagenesis. Most of these techniques involved standard directed evolution PCR methods.

4.3.1, Site directed mutagenesis (SDM) of G293E:

To characterise the novel NAAAR G293E enzyme discovered by Dr Royer, it was first necessary to mutate the WT sequence to incorporate the g874a DNA mutation required for this amino acid change (the original pTTQ18 NAAAR G293E plasmid cloned by Dr Sylvain Royer was not available). The WT glycine codon (GGG) was mutated to an *E. coli* glutamic acid codon (GAG) using the standard SDM approach with primers *G293E for* and *G293E rev*[249, 250]. The mutation is shown below in bold.

G293E for: 5'-GG TGC GGC G**A**G ATG ATC GAG ACC G -3'
G293E rev: 5'-CC GGT CTC GAT CAT C**T**C GCC GAC-3'

The mutagenesis was performed on the pTTQ18 NAAAR WT plasmid. Successful variants were compared to WT in the SET21 host and found to increase the rate of growth. The increase in growth rate was measured by colony size, judged by eye, on DMA plates containing 2.5 mM *N*-acetyl-D-methionine.

4.3.2, Saturation mutagenesis of G293:

A standard approach in directed evolution is to saturate a position highlighted by XL1-Red/epPCR with all other amino acid possibilities. This was performed at position G293 in pTTQ18 NAAAR WT using an NNK degenerate codon encoding all 20 proteinogenic amino acids[158].

4.3.2.1, PCR amplification of NAAAR G293x mutagenic library:

Saturation mutagenesis of G293 was performed on pTTQ18 NAAAR WT using primers *G293x for* and *G293x rev*. The degenerate NNK codon is shown below in bold[182].

G293x for: 5'-C CCG GTG TGG TGG TGC GGC **NNK** ATG-3'
G293x rev: 5'-CC GGT CTC GAT CAT **MNN** GCC G-3'

The PCR was performed with *Pfu turbo* and the product digested for 5 hours with *DpnI* at 37°C to remove WT template. *DpnI* was then inactivated by incubation at 80°C for 20 minutes in a PCR thermal cycler followed by gentle cooling to room temperature and centrifugation. Inactivation of DNA binding proteins was thought to help increase transformation efficiency.

4.3.2.2, Screening of NAAAR G293x mutagenic library:

The *DpnI* digested product was briefly centrifuged before 2 µL was used to transform 50 µL of electrocompetent SET21 cells. An ethanol precipitation of the digested product was tested, however this was found to decrease transformation efficiency by 50%. Transformed cells were incubated at 37°C for one hour with agitation and then washed three times in sterile ddH_2O to remove all exogenous sources of L-methionine. To screen the mutagenic library, 95% of the washed cells were spread on a DMA plate (2.5 mM *N*-acetyl-D-methionine, 30 µg mL^{-1} chloramphenicol, 100 µg mL^{-1} ampicillin, 1 mM IPTG) with the remaining 5% spread on an LB agar plate (30 µg mL^{-1} chloramphenicol, 100 µg mL^{-1} ampicillin). This LB agar plate was used to check cloning and transformation efficiency. After 24 hours incubation at 37°C, forty colonies were visible on the DMA plate. The four largest of these were re-streaked on replica DMA to isolate single colonies. After 24 hours growth, three of these streaks had grown better than the WT. The plasmid DNA from the three faster growing colonies was isolated and sequenced to reveal a single mutation common to all plasmid samples. The WT glycine codon had been

mutated from GGG to GAT, corresponding to aspartic acid[250]. As an NNK codon was used for saturation mutagenesis, this is the only aspartic acid codon available. A repeated round of screening under the same conditions generated only glutamic acid variants (from WT). All variants isolated from screening of G293 were found to contain either glutamic or aspartic acid, and SET21 hosts expressing these variant NAAARs grew faster than hosts expressing the WT. When the growth rates of the two variants were compared, it was observed that G293D grew slightly faster than G293E. An interesting observation from his process is that although the initial XL1-Red screen (performed by Dr Sylvain Royer) was able to generate NAAAR G293E as an enzyme with improved activity; it would have been highly unlikely for it to create a G293D variant within a single round of screening. The reason for this is that the chances of WT being mutated to G293E are roughly 1 in 3000 (as there are 1104 base pairs, with 3 possible base substitutions per base, only one of which, GGG to GAG, gives glutamic acid), whereas the chances of WT being mutated to G293D are roughly 1 in 4,500,000 (roughly 1 in 3000 to generate GAG, then 1 in 1500 to generate either GAT or GAC after that initial GAG mutation). These numbers show the importance of saturation mutagenesis as a means to ensure that as much of the useful protein sequence space is tested as possible during directed evolution.

4.3.2.3, Structural analysis of G293E/D mutations:

The G293E and G293D mutations share functional overlap as they both introduce a carboxylate group. This suggests that this functional group is of key importance in increasing NAAAR activity. In the WT NAAAR:*N*-succinyl-L-methionine crystal structure (PDB code: 1SJC) the peptidic carbonyl of G293 (G291 in the native form and PDB structure) forms a H-bonding interaction with the *N*-succinyl amide forming part of the *N*-acyl binding pocket[100]. This interaction does not appear to occur in the WT:*N*-acetyl-L-methionine structure (PDB code: 1SJA) as there is distance of 3.8 Å between these groups (**Figure 43**). The introduced glutamic and aspartic acid side groups may reposition the peptidic carbonyl to allow this interaction, or they may enter into the active site themselves and H-bond with the substrate amide. The conservation of the carboxylate group would suggest that this functionality is essential for whatever role these mutations play in increasing enzymatic activity.

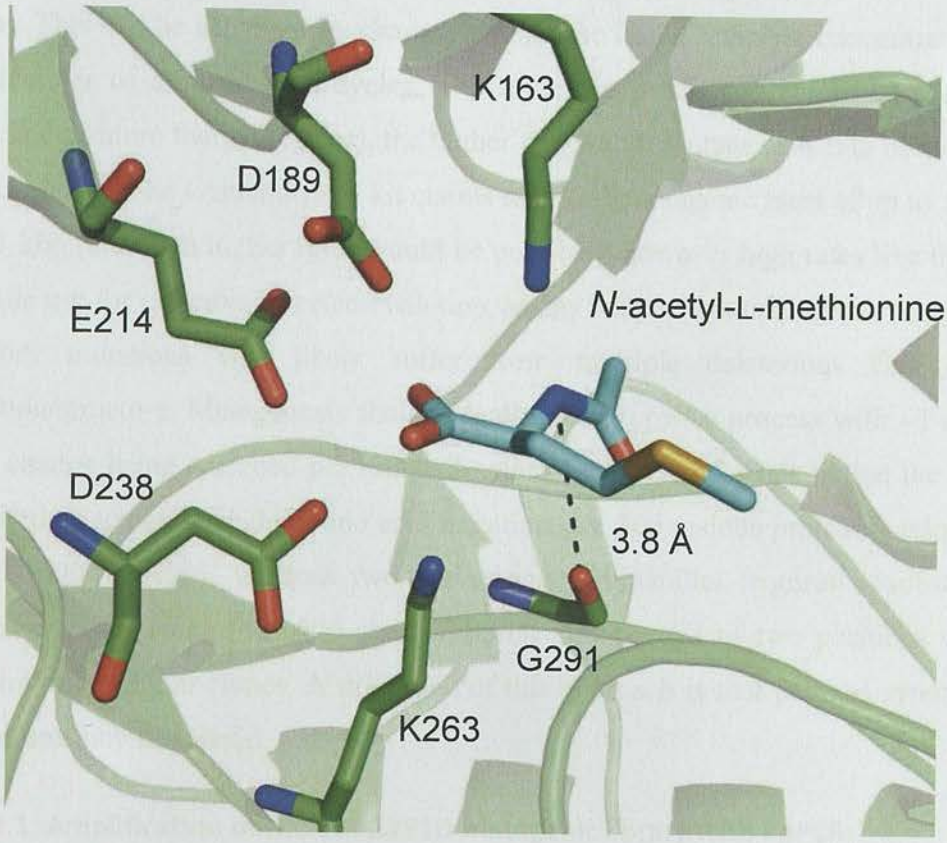


Figure 43. Location of G291 in NAAAR active site. Catalytic residues K163, K263, D189, E214, and D238 shown for clarity. *N*-acetyl-L-methionine shown in cyan. Figure generated from 1SJA[100].

4.3.3, Error prone PCR (epPCR) of NAAAR G291D:

After the success of saturation mutagenesis at position G293, a more random approach was taken for the next round of mutagenesis. At this stage screening was also switched from the pTTQ18/SET21 system to the improved pET20b/SET21(DE3). Error prone PCR can be accomplished by modification of a standard PCR, for instance, over inclusion of a single base in the dTNP mix, or swapping the usual *Taq* polymerase Mg^{2+} cofactor to Mn^{2+}, however these approaches tend to give a limited mutation spectrum that is not conductive to effective epPCR[161]. In order to achieve a varied mutational spectrum, a commercial kit (Genemorph II, Stratagene) was used. This kit uses a combination of *Taq* and a proprietary error prone polymerase that has been engineered to have the opposite mutational spectrum to that of *Taq* polymerase. As a result, this kit gives a mutagenic product with a balanced spectrum containing an equal distribution of transitions and transversions. Another useful feature of this commercial kit is that the level of mutagenesis is easily controlled by altering the total number of amplifications that

occur. This can be achieved by changing either the initial template concentration or the number of amplification cycles. The more the product is amplified (*i.e.* less template or more thermal cycles), the higher the overall mutagenesis rate of the final PCR product. The Genemorph II kit claims to allows mutagenic rates of up to 16 bps per 1 kbp (although higher rates would be possible), however high rates like this are of little use for effective directed evolution, as any enzyme containing this number of random mutations will likely suffer from multiple deleterious changes in function/structure. Mutagenesis should ideally be a stepwise process with ~1 amino acid change being screened per round. Another key reason for this is that the effort required to screen a single amino acid mutation in a 300 residue protein is relatively small (6000 clones), whereas two amino acids in parallel requires a substantial increase in screening (120,000 clones). Serial mutagenesis of two positions would require only 12,000 clones. A drawback of this approach is that pairs of synergistic mutations may be missed.

4.3.3.1, Amplification of NAAAR G291D mutagenic library with epPCR:

A mutagenic pET20b G291D NAAAR library was constructed using primers *pET20b Fm* and *pET20b Rm*. These primers terminate on the 3' end with either the ATG start codon or CTA stop codon limiting mutagenesis to residues 2 – 368 of the NAAAR protein. The start and stop codons are shown in bold and restriction sites underlined below.

pET20b Fm 5'-GTT TAA CTT TAA GAA GGA GAT ATA <u>CAT **ATG**</u>-3'
pET20b Rm 5'-GTG CTC GAG TGC GGC CGC <u>AAG CTT</u> **CTA**-3'

The NAAAR G291D gene was amplified using an initial template (*i.e.* the NAAAR gene) concentration of 100 ng and cycled 25 times. According to the manufacturer's specification, these conditions should generate between 0 - 4.5 mutations/kb. The amount of template is given as the amount of the actual gene, for instance, pET20b NAAAR G291D is roughly 5000 bp, of which 1100 bp are NAAAR template, therefore in a reaction containing 500 ng of total plasmid, 20% of this (100 ng) is the NAAAR template. Two µL of the epPCR product was then used as the template in a second, high fidelity PCR (*Pfu*) which is required to generate the >300 ng of PCR product needed for incorporation into the pET20b vector *via* a megaprimer of whole plasmid PCR (MEGAWHOP)[186, 197]. Also, as the epPCR product is relatively expensive to produce (£15/reaction), it makes sense to re-amplify this so it can be

used numerous times, and allow optimisation of the MEGAWHOP process. MEGAWHOP eliminates the need for a ligation that can severely limit the effectiveness of library construction. The MEGAWHOP works much like SDM; however, it uses a ~1100 bp gene as a mutagenic primer pair. Extension from the 3' end re-constructs the vector backbone once the gene has annealed at the correct position. The product of this process is a mutagenic nicked plasmid library that can be used to transform the SET21(DE3) host for selection. The conditions for the MEGAWHOP PCR were optimised to maximise the number of c.f.u. after SET21(DE3) electroporation. The optimum conditions were found to be 100 ng of pET20b NAAAR template and at least 300 ng of megaprimer. This was found to be a crucial parameter as MEGAWHOP efficiency dropped dramatically below 300 ng. The optimal number of thermal cycles was found to be 25 and increasing this caused no improvement in the number of c.f.u. After 35 cycles, the number of c.f.u. decreased. SET21(DE3) (at least $1x10^8$ c.f.u. per μg pUC18) electroporated with the MEGAWHOP PCR product under optimum conditions were estimated to contain approximately 10,000 c.f.u. (5% spread on an LB agar gave approximately 500 c.f.u.). To check that the mutagenic spectrum of this epPCR was unbiased, the *DpnI* digested MEGAWHOP product was used to transform DH5α and the plasmid DNA isolated. Four plasmids were sequenced to determine the mutagenic spectrum of the epPCR mutagenic library (**Table 3**).

Table 3. Variants sequenced from pET20b G291D NAAAR error prone PCR library.

Colony	*Mutations*
1	t629c (L210P)
2	g144t (silent), g642a (silent), a1061c (E353A)
3	c963g (E321D)
4	Parental template

Sequencing of these plasmids indicated that a good mutagenic spectrum had been achieved. The mutations corresponded to 40% transitions (C↔T, A↔G): 60% transversions (T↔AG, C↔AG, A↔TC, G↔CT) and CG→N (40%): AT→N (60%). The library contained on average 1.2 DNA mutations per sequenced gene (1104 bp), or 0.75 amino acid mutations per enzyme (368 residues).

4.3.3.2, Screening of pET20b G291D NAAAR mutagenic library:

To screen this library for improved variant NAAARs, five aliquots of SET21(DE3) were each electroporated with 2 μL of *DpnI* digested MEGAWHOP PCR product constructed from the above epPCR product. The cells were incubated at 37°C for 1 hour with agitation before being washed three times with 1 mL of sterile ddH_2O. The five aliquots were each resuspended in a final volume of 100 μL ddH_2O, of which 5 μL were spread on an LB agar plate (30 μg mL^{-1} chloramphenicol, 100 μg mL^{-1} ampicillin) and the remaining 95 μL was spread on a DMA plate (30 μg mL^{-1} chloramphenicol, 100 μg mL^{-1} ampicillin, 500 μM *N*-acetyl-D-methionine). After 16 hours growth at 37°C, the number of colonies on each of the five LB agar plates was approximately counted and this value used to judge the number of c.f.u. thought to have been screened. This was estimated to be in the region of around 35,000-40,000 c.f.u. in total from the five plates. After 48 hours growth, three colonies were judged by eye to have out grown their immediate neighbours. These three colonies were re-streaked onto a single replica DMA plate and a single colony (from each streak) from this plate used to inoculate 5 mL of LB broth (30 μg mL^{-1} chloramphenicol, 100 μg mL^{-1} ampicillin). The plasmid DNA was isolated from these three colonies and sequenced (**Table 4**).

Table 4. Mutations found within fast growing NAAAR colonies.

Colony	*Mutations*
1	c598t (P200S), t968a (F323Y)
2	t968a (F323Y)
3	a148t (M50L), g1065t (silent)

An initial assay of activity of these variant NAAARs was conducted using a liquid growth assay in Davis minimal broth (DMB). DMB is identical to DMA, but in liquid form, allowing for OD_{600} measurements to be read and a quantitative value for growth attributed to each SET21 colony/NAAAR variant. This approach was investigated as it proved very difficult to measure colony size accurately as they tended to be only 1-2 mm in diameter. The liquid growth assay was performed by taking 1 mL of an LB broth (100 μg mL^{-1} ampicillin) overnight culture of SET21(DE3) expressing each NAAAR variant and harvesting the cells with

centrifugation. The cell pellet was then washed three times with 0.22 μm filtered ddH_2O. The washed cell suspension was then diluted in DMB (30 μg mL^{-1} chloramphenicol, 100 μg mL^{-1} ampicillin, 500 μM *N*-acetyl-D-methionine) to an OD_{600} of ~0.15. The cultures (5 mL) were then incubated at 37°C with agitation and the increase in OD_{600} measured every hour. A standard of SET21(DE3) expressing NAAAR G291D was included along with a control containing no source of methionine to rule out non-methionine auxotroph contaminations. Typically, there was no OD_{600} change during the first 5 hours, after which differing rates of growth were observed for SET21(DE3) hosts expressing each NAAAR variant (**Table 5**). After 24 hours, all cultures had reached a maximum OD_{600} of ~0.9. Addition of further *N*-acetyl-D-methionine after this point did not result in an increase in OD_{600}.

4.3.3.3, Liquid growth assay of NAAAR G291D P200S F323Y, NAAAR G291D F323Y, and NAAAR G291D M50L variants:

To overcome the low sensitivity of the solid agar screen, a liquid screening assay was developed to quantify the rate of culture growth. The high number of false positives encountered highlighted the need for an improved quantitative screening method. Attempts were made at using a plate reader to carry out this work but it was found to be impossible to measure NAAAR dependent growth with the plate reader available. An effective setup was found that allowed reproducible growth using 5 mL cultures of minimal media incubated at 37°C on a flask shaker. The OD_{600} was measured using a spectrophotometer at set time points (**Table 5**).

Table 5. Comparison of growth characteristics for SET21 complemented with differing NAAAR variants.

NAAAR variant	*Initial OD_{600}*	*OD_{600} (8 hours later)*	*OD_{600} increase*
G291D	0.16 ± 0.01	0.18 ± 0.04	0.02
G291D M50L	0.14 ± 0.01	0.36 ± 0.01	**0.22**
G291D F323Y P200S	0.16 ± 0.01	0.24 ± 0.02	0.08
G291D F323Y	0.16 ± 0.01	0.49 ± 0.01	**0.33**

All three variant NAAARs were found to increase the growth rate of the SET21(DE3) host compared to the parental NAAAR G291D template. The largest

increase in growth came from the F323Y mutation, with M50L also providing a marked increase. Interestingly, the F323Y mutation was isolated in two independent colonies; the other also containing the P200S mutation. From the growth data, it would appear that F323Y is a beneficial mutation increasing SET21 growth, whereas P200S is deleterious. Given this, the P200S mutation was quickly discarded as an adverse mutation that had only been highlighted due to its occurrence in the same plasmid as the F323Y mutation.

4.3.3.4, Structural analysis of M50L, P200S, and F323Y mutations:

Analysis of the X-ray crystal structure of NAAAR WT complexed with *N*-acetyl-methionine (PDB code: 1SJA, substrate modelled as the L-enantiomer) gave an indication of what possible role the of M50L, P200S, and F323Y mutations may be playing[100]. Both M50L and F323Y are found in two separate substrate binding pockets; M50L appears to be located in the substrate side group binding pocket and F323Y appears to contribute to the *N*-acyl binding pocket along with G291D. The deleterious P200S was located on an outer α-helix, ~25 Å away from the substrate (α-carbon to α-carbon) and clearly played no direct role in the active site. The presence of proline in an α-helix often leads to kinks or bends, and in the WT X-ray crystal structure there is a distinct kink before P200[251]. This kink may help to position D189, a key catalytic residue involved in Mg^{2+} binding, within 2.5 Å of the metal cofactor. The removal of this kink may have pulled D189 into an unfavourable position causing a decrease in enzyme efficiency by altering the Mg^{2+} binding site (**Figure 44**).

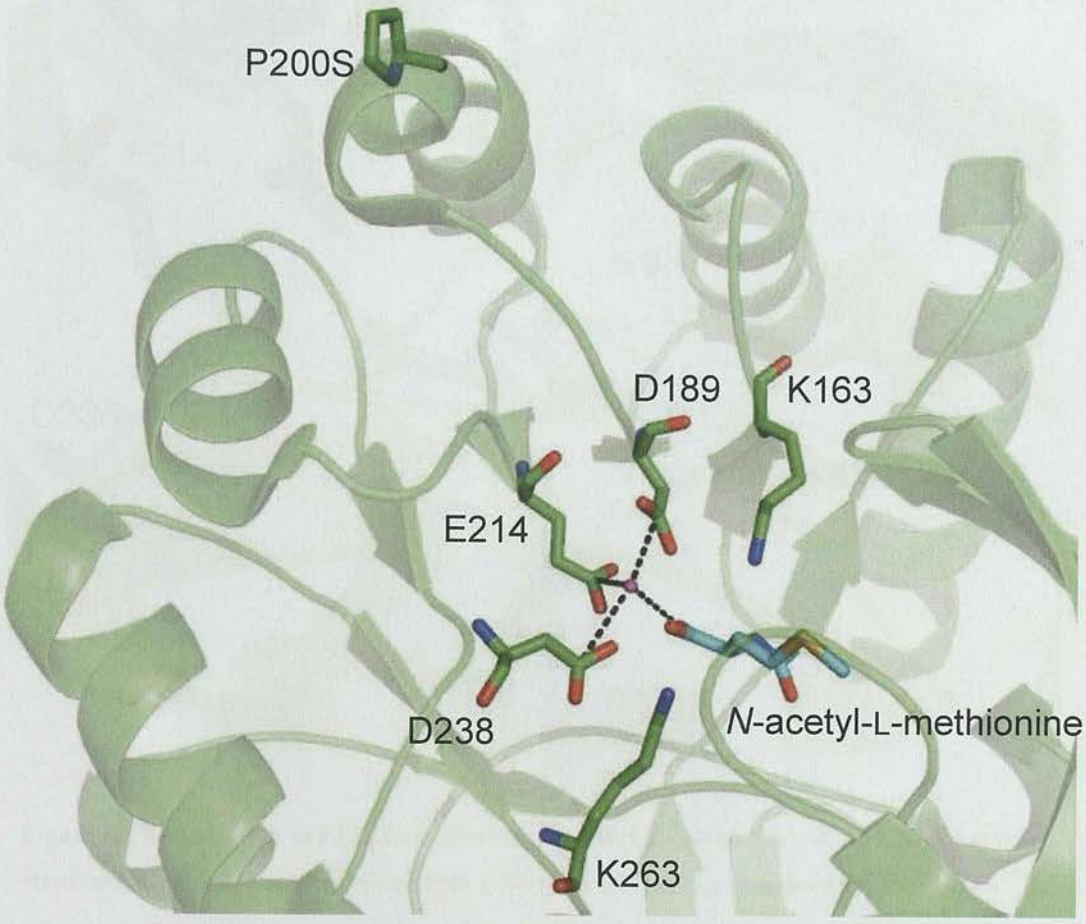


Figure 44. Location of P200 and its proximity to the Mg^{2+} binding site in NAAAR WT. NAAAR backbone shown as green ribbon. Figure generated from 1SJA[100].

Several encouraging observations were made for the F323Y mutation. Firstly it was present in two independent colonies that were not daughters from phenotypic outgrowth (due to the occurrence of P200S in only one of these). It appeared to increase the activity sufficiently in the presence of a likely deleterious mutation to allow an increase in SET21 growth. And thirdly, there was a clear chemical rational of how this mutation could influence substrate binding. The mutation from phenylalanine to tyrosine is a relatively small mutation structurally; however, there is a rather marked change in the chemistry once the introduced hydroxyl is considered. The hydroxyl can now provide an H-bond donor which could possibly interact with the substrate acetyl carbonyl. Given that in the WT structure the *para*-carbon of the phenylalanine ring resides only 5.8 Å from the C atom of this carbonyl, this is a feasible interaction (**Figure 45**).

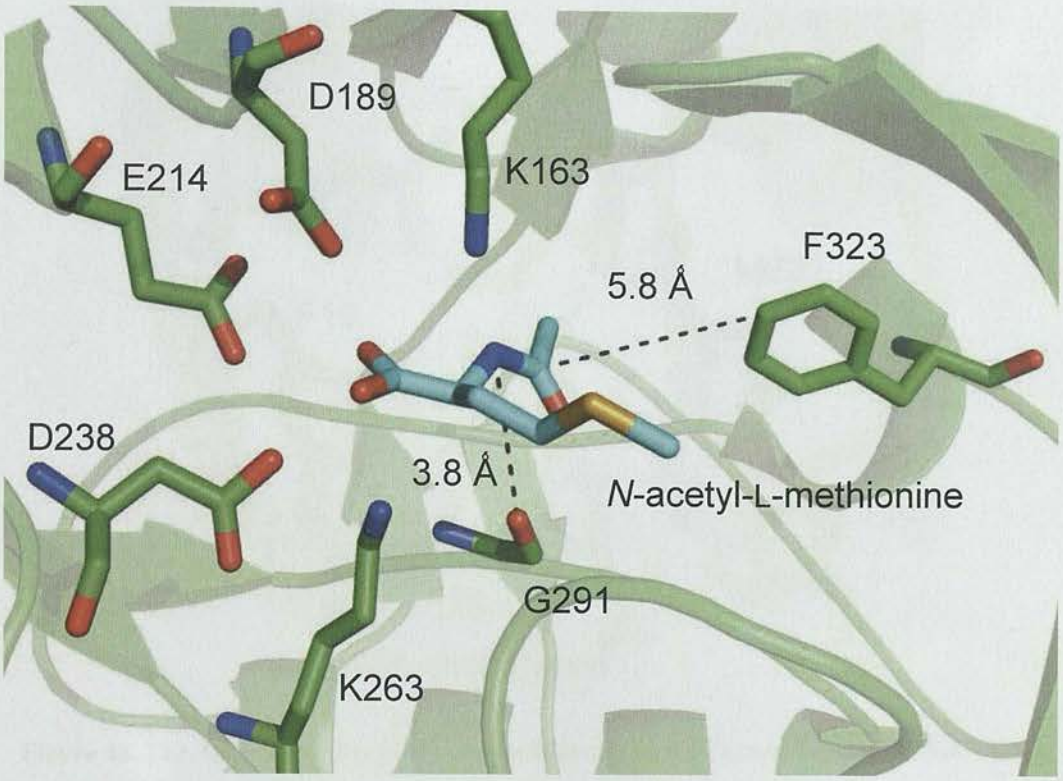


Figure 45. The position of F323 and substrate carbonyl group in NAAAR WT X-ray crystal structure. Catalytic residues, along with G291 also shown. Figure generated from 1SJA[100].

Although the F323Y mutation is novel to the *Am. Ts-1-60* NAAAR, in other bacteria such, as *G. kaustophilus,* the equivalent 323 position of NAAAR is already a tyrosine[252]. It may be that the next step in the natural evolution of *Am. Ts-1-60* NAAAR towards increased *N*-acyl amino acid racemase activity would have been the F323Y mutation.

M50L is located in the amino acid substrate binding pocket, close to the lid domain of the NAAAR monomer (**Figure 46**). The binding pocket consists of seven residues which surround the amino acid side group (F19, F23, Q26, M50, Y55, I292, and M293). The mutation of methionine to leucine is likely to affect the sterics of the binding pocket more than any chemical interactions as both amino acids are hydrophobic in nature. The distance between the terminal methyl groups of both M50 and substrate is only 3.1 Å. The M50L mutation may open up extra space allowing for better binding of the ethyl methyl sulphide side chain of the methionine substrate.

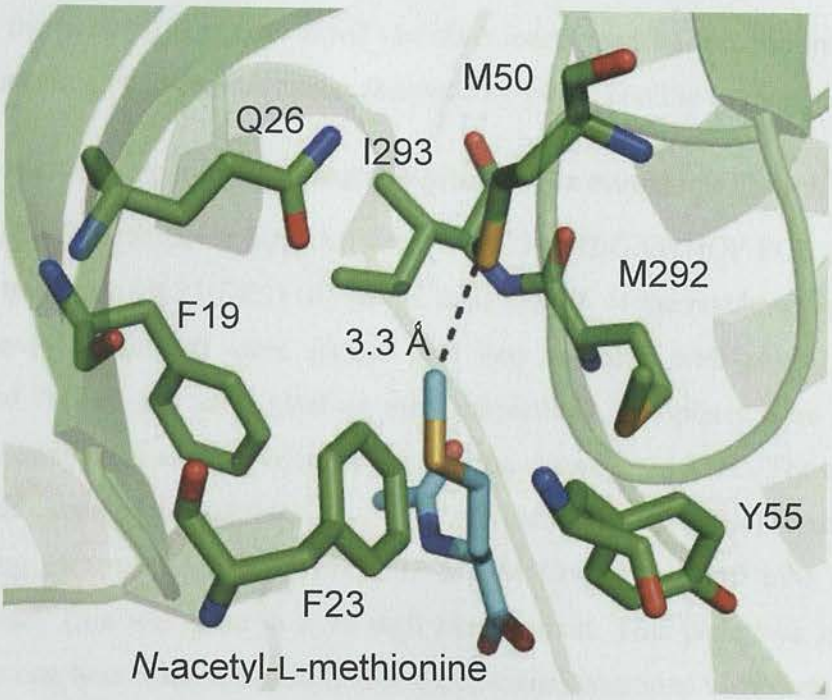


Figure 46. The *Am. Ts-1-60 NAAAR* amino acid substrate side group binding pocket. Figure generated from 1SJA[100].

4.3.4, Saturation mutagenesis of F323:

Saturation mutagenesis was carried out at position F323 to see if the F323Y mutation was most active amino acid at this position. This work was carried out using high-throughput HPLC screening to compare this process to the SET21 genetic selection.

4.3.4.1, PCR amplification of NAAAR F323x mutagenic library:

Saturation mutagenesis of F323 was performed by creating a megaprimer from pET20b NAAAR G291D using primers *pET20b Fm* and *F323x rev*. This megaprimer was incorporated into pET20b NAAAR G291D using a MEGAWHOP PCR. The degenerate MNN (reverse complement of NNK) codon shown in bold, and the NdeI restriction site underlined below.

pET20b Fm: 5'- GTT TAA CTT TAA GAA GGA GAT ATA CAT ATG -3'
F323x rev: 5'- GGT TTT GTG **MNN** CCG GTC C -3'

The PCR megaprimer product (~1100 bp) was gel purified and used in a MEGAWHOP PCR using pET20b NAAAR G291D as a template. The PCR was performed with *Pfu turbo* and the product digested for 5 hours with *DpnI* at 37°C to

remove the parental template. *DpnI* was then inactivated by incubation at 80°C for 20 minutes in a PCR thermal cycler followed by gentle cooling to room temperature.

4.3.4.2, High-throughput screening of NAAAR F323x mutagenic library:

The *DpnI* digested NAAAR G291D F323x MEGAWHOP PCR product was used to transform BL21(DE3) competent cells and 96 of the resulting transformants (from several hundred) were picked into two 48 deep well plates. Each well contained 700 µL LB broth (100 µg mL^{-1} ampicillin). The plates were sealed with gas permeable seals and incubated with agitation overnight at 37°C. The cells in each well were assayed by adding 10 µL of cell broth to 90 µL of reaction buffer containing 100 mM Tris-HCl (pH 8.0), 5 mM $CoCl_2$, and 150 mM *N*-acetyl-D-methionine. This was done in a 96 well plate format. This plate was incubated at 60°C for one hour with agitation before all biotransformations were terminated with a 1000 fold dilution with HPLC running buffer (75:25 0.01% TEAA:MeOH). The racemisation of *N*-acetyl-D-methionine was monitored by chiral HPLC. Separation of chiral molecules by HPLC can be achieved using two methods. Firstly, a chiral stationary phase (CSP) can be used to separate enantiomers by enantioselectively retaining one more than the other, thus imparting differing retention profiles on the two conformations, or by derivatisation of the analyte with a chiral auxiliary to form a pair of diastereomers that can be resolved by RP-HPLC on a non-chiral column[253]. In the case of primary amines such as amino acids, chiral derivatisation can be performed with *o*-phthaldialdehyde (OPA) and *N*-acetyl-L-cysteine (NAC)[254]. As the substrate/products for NAAAR are secondary amines, OPA/NAC derivatisation is not possible. To overcome this, a CSP in the form of a Chirobiotic™ T chiral column was used. This allowed the effective separation of both *N*-derivatised and free amino acids with no pre-column derivatisation required. The CSP used in this column is teicoplanin, an antibiotic glycopeptide (**Figure 47**)[255].

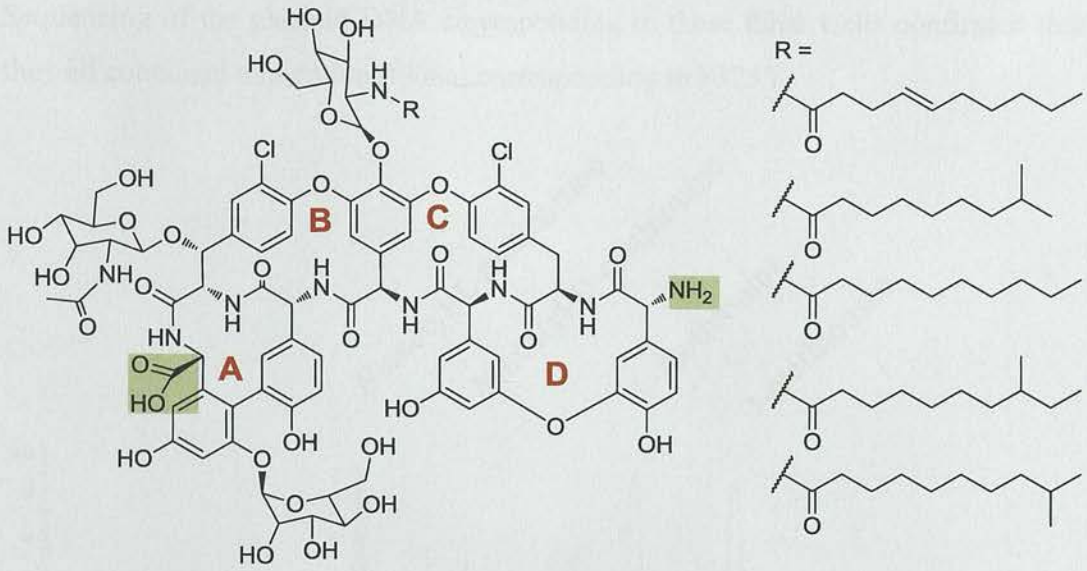


Figure 47. Teicoplanin, the CSP in Chirobiotic T chiral HPLC column.

The key structural features of this CSP that allow it to separate enantiomers are the ionic binding sites (green boxes), the alkyl chain attached to the β-D-glucosamine ring, and the four inclusion pockets (**A**, **B**, **C**, and **D**). Along with these, there are also π-acceptor/donator and H-bonding sites. The chiral environment of these functional groups provides the enantioselectivity required for separation. A literature report showing the use of this column to separate *N*-acetyl-amino acids was used as a starting point for optimised analysis conditions[256]. The final conditions varied very little from those published, with only a 5% increase in methanol required to provide a 10 minute HPLC run, monitored at 210 nm, that was sufficient to resolve both enantiomers of *N*-acetyl-methionine and also non-derivatised methionine in a single HPLC run. The chiral separation is attained by selective retention of the D-enantiomer of both molecules. The mobile phase used for separation was a 10 minute isocratic elution of 75:25 0.01% TEAA:MeOH which separated *N*-acetyl-L-methionine and *N*-acetyl-D-methionine by over 2 minutes. The free amino acids also eluted well separated under these conditions (**Figure 48**). Using automated injection and a 10 minute HPLC run, screening of a 96 well plate was completed 18 hours.

Screening of pET20b G291D F323x libraries was carried out by auto-sampling 20 μL of each 1000 fold diluted biotransformation for HPLC injection. The decrease in e.e. was used to measure NAAAR activity. Of the 96 wells tested, three had clearly higher levels of racemisation than the rest as judged by the HPLC traces.

Sequencing of the plasmid DNA corresponding to these three wells confirmed that they all contained a mutation (t968a) corresponding to F323Y.

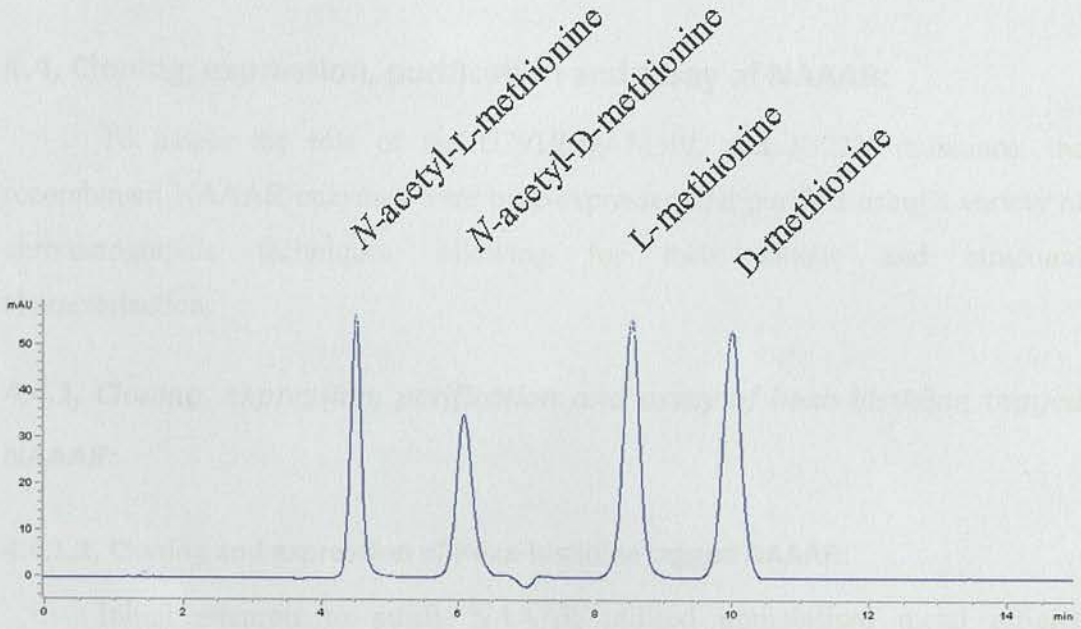


Figure 48. Injection of *N*-acetyl-DL-methionine and DL-methionine on a Chirobiotic™ T column monitored at 210 nm. Eluted with an isocratic elution of 75:25 (0.01% TEAA: MeOH) at 0.5 ml min^{-1}. Retention times are 4.5 minutes for *N*-acetyl-L-methionine, 6.1 minutes for *N*-acetyl-D-methionine, 8.5 minutes for L-methionine, and 10.0 minutes for D-methionine.

High-throughput screening of mutagenic NAAAR libraries proved to be an effective method for screening. Although the number of colonies screened (96) is considerably lower than the thousands that could be screened using the SET21(DE3) selection system, the output of actual enzymatic performance shown in an HPLC trace makes it much easier to identify positive hits. Although HPLC screening did not highlight a novel mutation, it did select G291D F323Y from a G291D template indicating that novel mutations that increase activity can be found using this method. HPLC screening gave a much higher sensitivity than the SET21(DE3) selection system which although was used successfully in this project, was only able to identify two mutations from screening of several thousand variant enzymes. A targeted mutagenesis approach in which the amino acids within the active site are subjected to iterative saturation mutagenesis (CASTing) and screened using high-throughput methods would possibly highlight novel mutations missed by the SET21 selection strain[160, 188, 190, 191]. Those mutation missed by the SET21 selection screen are those that increase activity, but do not increase activity enough to cause a substantial change in the SET21(DE3) phenotype. The difference in agar plate

growth between SET21(DE3) expressing NAAAR WT and those expressing NAAAR G291D F323Y is relatively small, making positive 'hit' identification difficult.

4.4, Cloning, expression, purification and assay of NAAAR:

To assess the role of the G291E/D, M50L and F323Y mutations, the recombinant NAAAR enzymes were over-expressed and purified using a variety of chromatographic techniques, allowing for their kinetic and structural characterisation.

4.4.1, Cloning, expression, purification and assay of hexa-histidine tagged NAAAR:

4.4.1.1, Cloning and expression of Hexa-histidine tagged NAAAR:

Initial attempts to purify NAAAR utilised immobilised metal affinity chromatography (IMAC) and expression of NAAAR from pET28a to incorporate an *N*-terminal hexa-histidine tag. The cloning of this plasmid was performed by amplification of NAAAR WT gene from pTTQ18 NAAAR WT using primers *NAAAR NdeI F* that encoded a 5'- NdeI site and *NAAAR HindIII R* that encoded a 3'-HindIII site. As the native start codon for *Am. Ts-1-60* NAAAR (GTG) is retained in the pTTQ18 plasmid, this was mutated to ATG to allow more favourable expression in *E. coli* systems. The restriction sites are shown in bold and the mutation underlined below.

NAAAR NdeI F 5'-**CAT <u>A</u>TG** AAA CTC AGC GGT GTG GAA CTG-3'
NAAAR HindIII R 5'-**AAG CTT** CTA CGA ACC GAT CCA-3'

The PCR product containing the NAAAR gene was gel extracted and ligated into the pGEM T-easy cloning vector. From here, the plasmid was digested with *NdeI* and *HindIII* before ligation into pET28a digested with the same enzymes. Sequencing confirmed the fidelity of the pET28a NAAAR WT clone (6415 bp). To allow expression and characterisation of NAAAR G291E, the NAAAR G293E gene was cloned from pTTQ18 NAAAR G293E by the same means. This generated pET28a NAAAR G291E. As with pET20b mentioned earlier, on going from pTTQ18 to pET28a, G293E becomes G291E. NAAAR WT and NAAAR G291E were over-expressed in BL21(DE3) for 5 hours with 0.1 mM IPTG at 37°C. Expression was

carried out in 5 L of LB broth (30 μg mL^{-1} kanamycin). Cells were harvested with centrifugation (*4,000 g*, 20 mins, 4°C) and stored at -20°C until needed.

4.4.1.2, Purification of hexa-histidine tagged NAAAR:

Purification of NAAAR was performed by first re-suspending a BL21(DE3) cell pellet expressing NAAAR in 10 mL of lysis buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 20 mM imidazole, 1 complete EDTA free mini protease inhibitor tablet (Roche)). Once fully re-suspended, the cells were lysed by sonication and the CFE clarified by centrifugation (*12,000* g, 60 mins, 4°C). The CFE was then filtered and incubated for 1 hour with pre-equilibrated Ni:NTA resin (Invitrogen) at 4°C. This resin was then loaded into a 1 mL polypropylene column (Qiagen) and washed with 40 mL of elution buffer (50 mM Tris-HCl, (pH 7.5), 100 mM NaCl, 20 mM imidazole) followed by a stepwise gradient of increasing imidazole concentration. The resin was washed with 20 mL of 20 mM, 10 mL of 150 mM and finally, 10 mL of 250 mM imidazole. The purification process, which was identical for both WT and G291E NAAAR, was monitored by SDS-PAGE analysis (**Figure 49**). This procedure was tested on a pre-packed HisTrap™ FF 1 mL Ni:Sepharose column (GE Healthcare) but the final product contained several contaminating proteins. All purification steps were performed on ice or at 4°C.

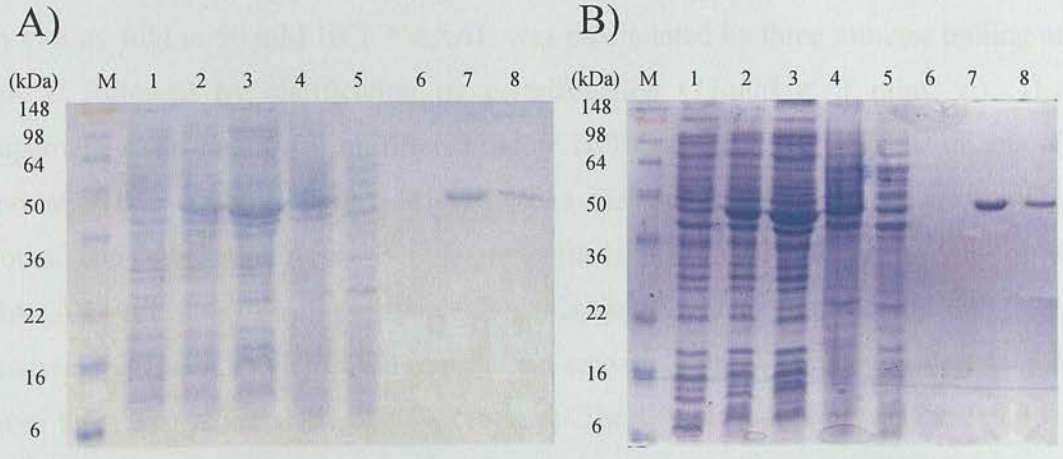


Figure 49. SDS-PAGE analysis of hexa-histidine tagged (A) NAAAR WT and (B) NAAAR G291E IMAC purification. M = See Blue Plus2 Marker, 1 = before induction, 2 = after induction, 3 = insoluble cell fraction, 4 = cell free extract, 5 = flow through, 6 = 20 mM imidazole wash, 7 = 100 mM imidazole elution, 8 = 250 mM imidazole elution.

Although a large amount of the protein was found in the insoluble cell fraction, there was still sufficient NAAAR in the CFE to allow for an effective purification. Due to

the high purity of the product (as judged from SDS-PAGE analysis), no further purification was performed. This process typically yielded 10 mg L^{-1} growth media for both recombinant NAAARs. The 100 mM and 250 mM imidazole fractions containing NAAAR were pooled and dialysed three times against 1 L of 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, after which, NAAAR concentration was measured by the absorbance at 280 nm (Abs_{280}). The following parameters, generated by Vector NTi 10 (Invitrogen), were used in calculating concentrations; mass of WT = 41,678 Da, mass of G291E variant = 41,750 Da, ε WT & G291E variant = 33,600 M^{-1} cm^{-1}[257]. NAAAR was found to be stable for 3-4 weeks when stored in 20% glycerol at -20°C.

4.4.1.3, *In vitro* assay of hexa-histidine tagged NAAAR:

To compare the activities of NAAAR WT and NAAAR G291E, kinetic characterisation of these two proteins was carried out using chiral HPLC. Enzyme assays were performed by incubating purified histidine tagged NAAAR with 100 µM - 20 mM of *N*-acetyl-D-methionine and *N*-acetyl-L-methionine. The reaction was performed at pH 7.5 in 50 mM Tris-HCl with 5 mM $CoCl_2$ as a cofactor. 100 µL of enzyme (0.1 mg) was incubated for three minutes at 37°C in a heat block with 900 µL reaction mixture containing buffer and substrate. All components were preheated to 37°C before mixing. After three minutes, the reaction was terminated and diluted by twenty fold in 50 mM HCl. NAAAR was precipitated by three minutes boiling at 100°C followed by clarification by centrifugation (*11,000* g, 3 mins, rt). The supernatant was then 0.45 µm filtered before HPLC analysis. To limit the effects of formation of methionine sulfoxide and sulfone, the reactions were analysed within 24 hours. The termination step was attempted with 10 mM EDTA, a known inhibitor of NAAAR and other metallo-enzymes, but this was found to co-elute and mask the analyte species in the HPLC traces[95]. The enzyme kinetic parameters determined from these assays are shown below (**Table 6**). The G291E mutation was expected to increase enzyme turnover, however, the results suggested that the G291E variant NAAAR is less active than the WT. This was a surprising observation as SET21 growth studies clearly showed an increase in SET21 growth when the host was expressing NAAAR G293E compared to the WT. It should be noted that although the K_m is lower for the G291E variant, the WT was more active at all substrate concentrations tested (100 µM - 20 mM).

Table 6. ***Am. Ts-1-60*** **NAAAR kinetic parameters for histidine tagged NAAAR WT and NAAAR G291E.**

Variant	*N-acetyl-L-methionine*		*N-acetyl-D-methionine*	
	K_m *(mM)*	k_{cat} (s^{-1})	K_m *(mM)*	k_{cat} (s^{-1})
WT	3.2 ± 0.40	2.15 ± 0.13	1.78 ± 0.25	1.36 ±0.08
G291E	2.64 ± 0.17	1.39 ± 0.06	1.42 ± 0.2	0.80 ±0.07

It was thought that as the GGG glycine codon is a rather unfavourable codon for *E. coli* expression and the GAG glutamic acid codon is more favourable, that an increase in protein expression could be accountable for the increase in SET21 growth. This was tested by mutating the rare GGG codon to another, more favourable glycine GGC codon[250]. When this DNA variant NAAAR was compared to WT and G293E, no increase in SET21 growth was observed ruling out the possibility that the increase in expression was due to replacement of a rare codon.

The observed kinetics were slightly lower than those published by Gerlt *et al* under similar conditions (**Table 2**) and inhibition was observed above 8 mM of both substrates compared to the 50 mM that is reported in the literature[95, 101]. A literature review of enolase enzymes revealed that their β-barrel structure has been noted to suffer from mis-folding when expressed with a histidine tag[258]. With this in mind, a new expression and purification process was developed based on the published method by Gerlt *et al* which allowed expression of native NAAAR from pET20b[100, 101].

4.4.2, Cloning, expression, purification and assay of native NAAAR:

4.4.2.1, Cloning and expression of native NAAAR:

To allow expression of native NAAAR, the pET20b expression vector was used. This vector also allowed for a "gentle" expression without any IPTG as there is no *lac* operator downstream of the T7 promoter. This plasmid had already been cloned for use in the SET21(DE3) selection process. The use of a single plasmid for both SET21(DE3) selection and protein expression would remove the need for an intermediate cloning step between these two processes. Expression from pET20b was found to be excellent in a BL21(DE3) host with no IPTG added for induction.

Expression was initiated by picking a single BL21(DE3) pET20b NAAAR colony into 500 mL LB broth (100 μg mL^{-1} ampicillin). These were left to grow at 37°C for 24 hours with agitation (**Figure 50**).

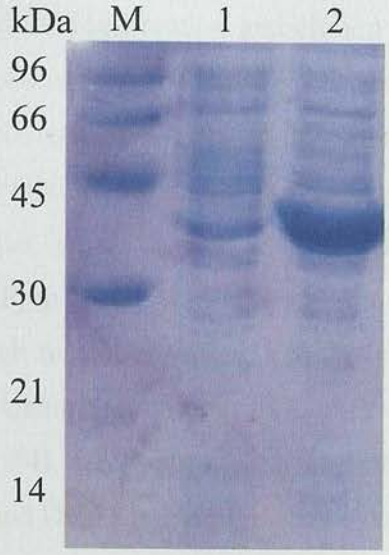


Figure 50. Leaky BL21(DE3) expression from pET20b NAAAR WT over 24 hours. M = L.M.W marker, 1 = BL21(DE3) pET20b after 24 hours, 2 = BL21(DE3) pET20b NAAAR WT after 24 hours.

Typically, expression was started at 9 am, and cells harvested (*4,000* g, 20 mins, 4°C) 24 hours later allowing for NAAAR purification to be performed with fresh cells that had not been frozen. If required, cell pellets were stored at -20°C.

4.4.2.2, Purification of native NAAAR:

NAAAR purification was performed using a combination of ion exchange (IEX) and size exclusion chromatography (SEC). This method was identical for all NAAAR variants and the WT. As the isoelectric point (pI) of NAAAR is 4.97, the monomer has a charge of -13.1 at pH 8.0, hence, anion exchange should allow NAAAR to be separated from the less negatively charged background BL21(DE3) proteins[259]. It was observed during the anion exchange optimisation process that if no NaCl was included in the lysis buffer, NAAAR was insoluble. Therefore all lysis buffers contained 100 mM NaCl. Three anion exchange columns were tested for their capability as an initial purification step; these were a Hi-Load Sepharose Q 26/10 (50 mL), a Resource Q (6 mL), and a HiPrep FF Q 16/10 (20 mL). All columns were equilibrated with 50 mM Tris-HCl (pH 8.0) 100 mM NaCl and elution performed with the same buffer containing 1 M NaCl. Although all three of these columns

worked well as an initial purification step, there were inherent difficulties with two of the columns. The Sepharose Q required dilution of the CFE from 10 mL to 100 mL and ten separate loadings *via* a superloop – this step alone took over two hours to perform – compared to a single injection for the other two columns. This, and the several hours (>12) required for equilibration and elution made using this column a long and labour intensive purification. The Resource Q was found to be a very quick method for purification of NAAAR, however, as this column has been discontinued by the manufacturer, it was rejected as should it ever stop functioning, it would be impossible to replace. This decision was made with X-ray crystallography particularly in mind. The HiPrep FF Q 16/10 was found to be the best column for the purification process with high resolution giving a highly pure NAAAR protein after only one purification step. Using this column, NAAAR was found to elute at 300 mM NaCl (fractions 27 to 34), well separated from both the background *E. coli* proteins (fractions 6 to 25) and DNA (fractions 40 onwards) (**Figure 51**).

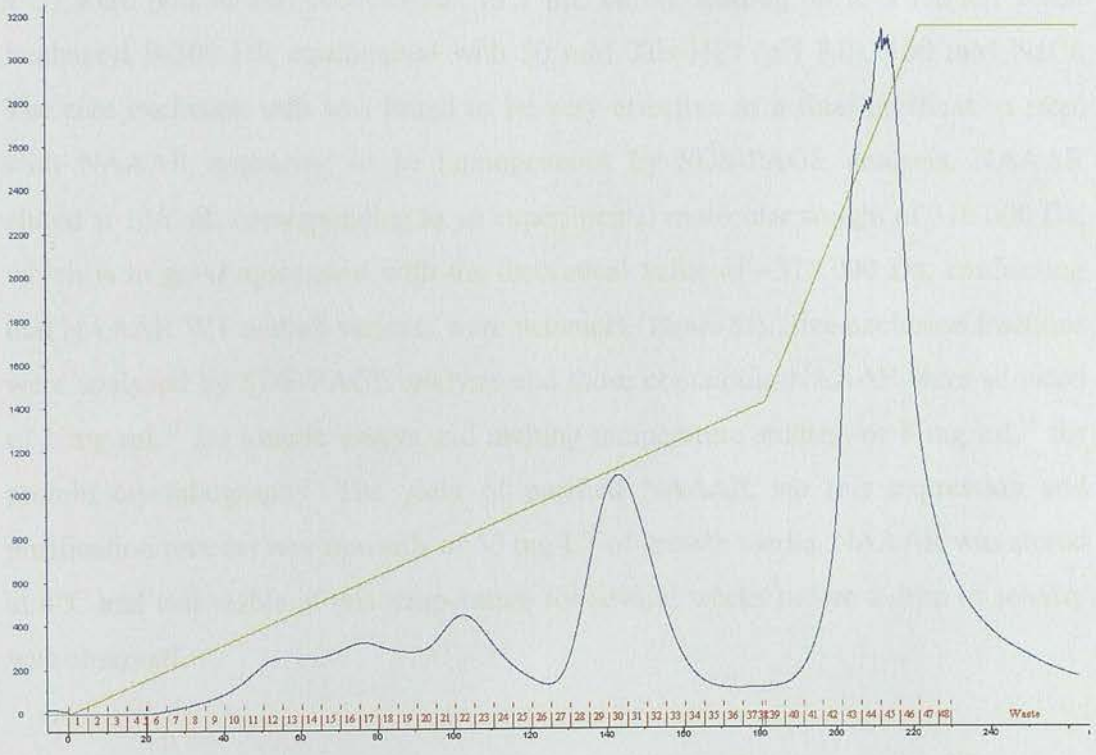


Figure 51. HiPrep FF Q 16/10 chromatogram trace of absorbance at 280 nm. NaCl gradient shown as light green trace. NAAAR elutes as a symmetrical peak between fractions 27 and 34. Column gradient was 0 to 45% high salt over 9 column volumes (cv), followed by 45 to 100% high salt over 3 cv.

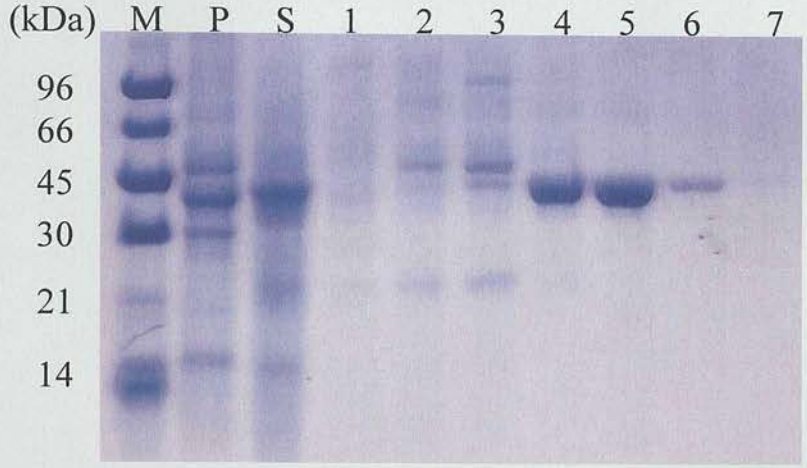


Figure 52. SDS analysis of HiPrep FF Q 16/10 fractions. NAAAR elution shown in lanes 4 + 5. M = L.M.W marker, P = insoluble pellet, S = soluble fraction, 1 = fractions 11 – 15, 2 = fractions 16 – 20, 3 = fractions 21 – 25, 4 = fractions 26 – 30, 5 = fractions 31 – 35, 6 = fractions 36 – 40, 7 = fractions 41 – 48.

Fractions found to be containing a high purity of NAAAR (**Figure 52**, lanes 4 + 5) were pooled and concentrated to 1 mL before loading on to a HiPrep 26/60 Sephacryl S-300 HR equilibrated with 50 mM Tris-HCl (pH 8.0), 100 mM NaCl. The size exclusion step was found to be very effective as a final purification step, with NAAAR appearing to be homogeneous by SDS-PAGE analysis. NAAAR eluted at 136 mL corresponding to an experimental molecular weight of 310,000 Da, which is in good agreement with the theoretical value of ~315,000 Da, confirming that NAAAR WT and all variants were octamers (**Figure 53**). Size exclusion fractions were analysed by SDS-PAGE analysis and those containing NAAAR were adjusted to 1 mg mL^{-1} for kinetic assays and melting temperature studies, or 8 mg mL^{-1} for protein crystallography. The yield of purified NAAAR *via* this expression and purification process was upwards of 50 mg L^{-1} of growth media. NAAAR was stored at 4°C and was stable at this temperature for several weeks before a drop in activity was observed.

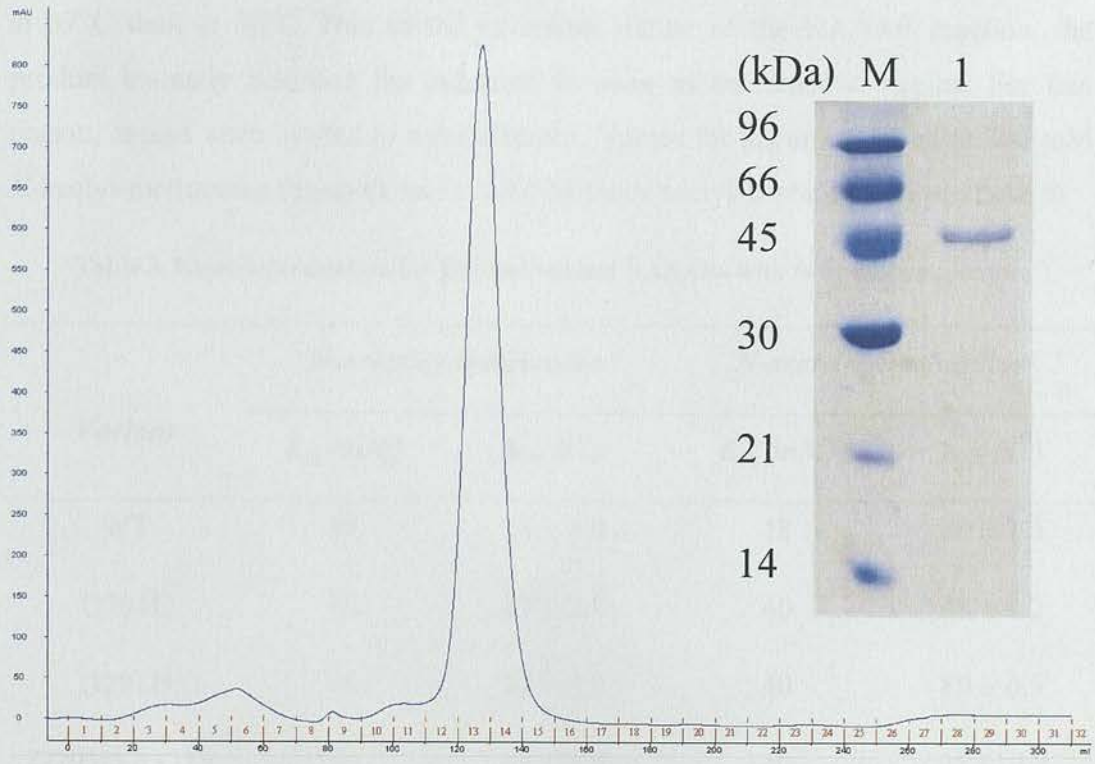


Figure 53. HR S-300 chromatogram trace of absorbance at 280 nm. SDS analysis of fractions 13 + 14 inset showing homogenous NAAAR. M = L.M.W marker, 1 = fractions 13 + 14 from S-300 purification.

Protein masses were determined using ESI FT ICR MS coupled to a RP-HPLC (SIRCAMS, The University of Edinburgh) and found to be within 1-2 Da of the expected values (**Table 7**).

Table 7. Theoretical and experimental masses (average neutral mass) for variant NAAARs.

Variant	*Theoretical mass (Da)*	*Experimental mass (Da)*
WT	39,407	39,405
G291E	39,479	39,478
G291D	39,465	39,463
G291D F323Y	39,481	39,480

4.4.2.3, *In vitro* assay of native NAAAR:

Native NAAAR was assayed in a similar manner to the histidine tagged protein, using chiral HPLC with a Chirobiotic™ T column for analysis. Assays were performed with various concentrations of *N*-acetyl-methionine (both enantiomers) and *N*-acetyl-D-phenylalanine at 60°C. Activity was found to be roughly 10% lower

at 37°C than at 60°C. Due to the reversible nature of the NAAAR reaction, the product instantly becomes the substrate as soon as the reaction begins. For this reason, assays were limited to three minutes. Values for k_{cat} are reported at 300 mM *N*-acetyl-methionine (**Table 8**), and at 10 mM for *N*-acetyl-D-phenylalanine (**Table 9**).

Table 8. Kinetic parameters for WT and variant NAAARs with *N*-acetyl-methionine.

Variant	*N-acetyl-D-methionine*		*N-acetyl-L-methionine*	
	K_m *(mM)*	k_{cat} (s^{-1})	K_m *(mM)*	k_{cat} (s^{-1})
WT	40	14 ± 1.9	18	20 ± 1.5
G291E	40	27 ± 2.8	40	48 ± 4.2
G291D	40	54 ± 3.8	40	80 ± 6.9
G291D F323Y	50	67 ± 5.5	40	95 ±10.7

As expected, based on the increase in growth rates observed with SET21 expressing these NAAARs, enzyme activity increases on going from WT → G291E → G291D → G291D F323Y. The native NAAAR G291E enzyme was now found to be more active than WT as suggested by growth studies. With the most active NAAAR G291D F323Y variant, k_{cat} has increased by ~4.7 fold after three rounds of evolution for both enantiomers indicating that the mutations are not affecting NAAAR enantioselectivity. No decrease in K_m was observed after each round of evolution suggesting the G291E/D and F323Y mutations are having very little, if any, effect on substrate binding.

NAAAR G291D M50L was purified and assayed as above but the M50L mutation was found to increase activity specifically for *N*-acetyl-DL-methionine. This mutation did not increase activity towards *N*-acetyl-D-phenylalanine above that of the WT, and caused a decrease relative to NAAAR G291D. The combination of G291D and F323Y was found to nearly double the activity of NAAAR towards *N*-acetyl-D-phenylalanine (**Table 9**). Values of k_{cat} for *N*-acetyl-D-phenylalanine are reported with 10 mM substrate due to substrate solubility issues.

Table 9. k_{cat} values for NAAAR WT, G291D, G291D M50L, and G291D F323Y with *N*-acetyl-D-phenylalanine.

Variant	*k_{cat} (s^{-1}) for N-acetyl-D-phenylalanine*
WT	12.32 ± 0.97
G291D	15.08 ± 0.8
G291D M50L	13.05 ± 1.3
G291D F323Y	21.90 ± 1.56

4.4.2.4, Substrate/product inhibition of NAAAR:

It is important that industrial biocatalysts have a high tolerance to substrate and product concentration to allow high productivity. In the case of NAAAR, this is essential as the acylases to which it will be coupled operate upwards of 500 mM substrate. The literature suggests that *Am. Ts-1-60* NAAAR, and other *Amycolaptosis* NAAARs suffer from substrate inhibition above 50 mM[95, 260]. This would seriously hinder their application to industrial processes[89]. Inhibition was confirmed by early *in vitro* assays in which no activity was observed above 50 mM *N*-acetyl-methionine with purified NAAAR. The reason for the inhibition was discovered during testing of a potential NAAAR variant at 90 mM *N*-acetyl-D-methionine. All potential variants were tested at this concentration to evaluate their impact on substrate inhibition. However during this particular experiment, six fold more HEPES buffer was added to this substrate concentration than usual. As a result, activity was observed at 90 mM for the first time. The drop in activity in the non-excess buffered reactions as they approached 50 mM confirmed that the observed substrate inhibition was due to the lack of buffering at these increasing acidic substrate concentrations (**Figure 54**). Further testing showed that NAAAR WT was active up to tested levels of 300 mM *N*-acetyl-methionine in 100 mM Tris-HCl (pH 8.0) if the substrate was pH adjusted to 8.0 before addition to the assay.

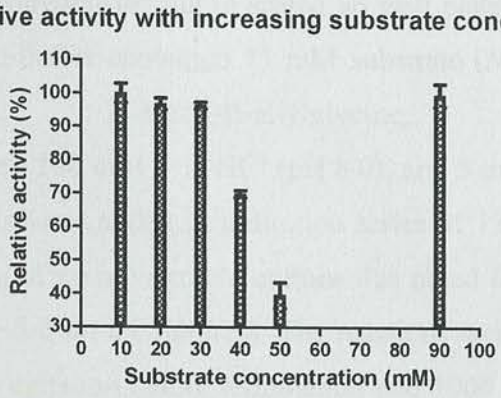


Figure 54. Inhibition of NAAAR above 50 mM substrate concentration with 50 mM HEPES. Activity restored at 90 mM with 300 mM HEPES. 100% activity plotted relative to 10 mM substrate.

4.4.2.3, *In vivo* assay of native NAAAR as a whole cell biocatalyst:

In vivo NAAAR assays were also carried out by expressing different NAAAR enzymes from pET20b in BL21(DE3) hosts, and then assaying these cells directly for NAAAR activity. This assay was performed to mimic the industrial conditions that NAAAR is likely to be used in, *i.e.* to ensure that NAAAR can be used in a whole cell biotransformation or as a partially purified enzyme. At this stage, potential industrial substrates (*N*-acetyl-allylglycine and *N*-acetyl-(4-fluoro)phenylglycine were introduced as it was hoped that any evolved NAAAR would be applicable to a range of high value, synthetically useful amino acids and not solely methionine. To test this, all four NAAAR variants were expressed for 24 hours in BL21(DE3) grown at 37°C in LB broth (100 μg mL^{-1} ampicillin) with no IPTG (**Figure 55**).

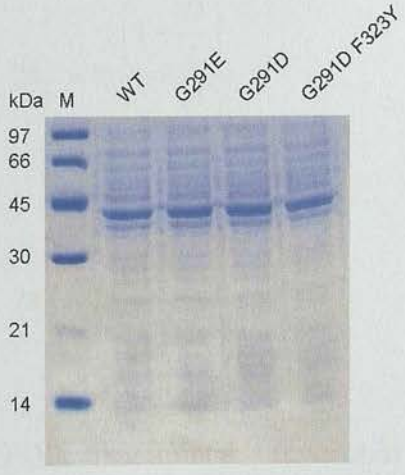


Figure 55. BL21(DE3) hosts used for whole cell biotransformations expressing NAAAR WT, G291E, G291D, and G291D F323Y.

Biotransformations were carried out in sealed 96 well plates with a reaction volume of 100 µL. Reaction buffer contained 75 mM substrate (*N*-acetyl-L-methionine, *N*-acetyl-D-methionine, *N*-acetyl-D-allylglycine, or *N*-acetyl-D-(4-flouro)phenylglycine), 100 mM Tris-HCl (pH 8.0), and 5 mM $CoCl_2$. To this buffer, 10 µL of whole cells were added as a dilution series of 1.0 (no dilution), 0.75, 0.5 and 0.25. The OD_{600} of each overnight culture was noted for calculation of specific activities (typically ~5.0 for all cultures). The reactions were allowed to proceed for 1 hour at 60°C with agitation before termination and 1000 fold dilution with HPLC running buffer (75:25 0.01% TEAA:MeOH). The conversion of the single enantiomer to a corresponding racemate was monitored using chiral HPLC. By using the conversion values for each dilution, a rate constant was generated for each variant with each substrate. The rate constants for amino acid racemisation (k_L and k_D) have been studied and modelled as a means of determining the age of fossils (**Figure 56**)[261, 262]. Rate constants k_L and k_D can be accurately modelled with first order reaction kinetics (FOK) as *Eqn (1)* for racemisation of L-amino acids, and *Eqn (2)* for the racemisation of D-amino acids where D = proportion of D-enantiomer, and L = proportion of L-enantiomer:

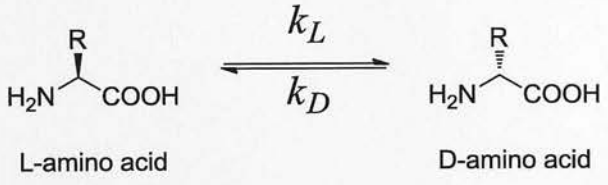


Figure 56. Rate constants k_L and k_D for the racemisation of amino acids.

Eqn (1),
$$2k_L = \ln \frac{\left(1+\frac{D}{L}\right)}{\left(1-\frac{D}{L}\right)}$$

Eqn (2),
$$2k_D = \ln \frac{\left(1+\frac{L}{D}\right)}{\left(1-\frac{L}{D}\right)}$$

A linear regression of the FOK transformed D/L or L/D measurements against enzyme concentration or time can be used to calculate k_L or k_D (hour^{-1}) by *Eqn (3)*:

Eqn (3),

$$k_{D/L} = \frac{slope}{2}$$

This assumes that $k_L = k_D$, which is not case for NAAAR, but as the transformation of the data gave good linearity with R^2 generally >0.95, it was taken to be an accurate model for NAAAR racemisation. Enzyme dilutions were used as effective time points, *i.e.* a 0.25 dilution over an hour = 15 minutes with no dilution. This model was used to calculate k_L and k_D for whole cell biocatalysts expressing variant NAAARs with *N*-acetyl-L-methionine, *N*-acetyl-D-methionine, *N*-acetyl-D-allylglycine, and *N*-acetyl-D-(4-fluoro)phenylglycine. The results are listed below for each of these substrates (**Figure 57** to **Figure 60** and **Table 10** to **Table 13**). Specific activities (μmoles $hour^{-1}$ OD_{600}^{-1}) where calculated using *Eqn (4)*:

Eqn (4),

$$specific\ activity = \frac{\left((total\ \mu moles\ of\ substrate)k_{D/L}\right)}{OD_{600}}$$

***N*-acetyl-L-methionine:**

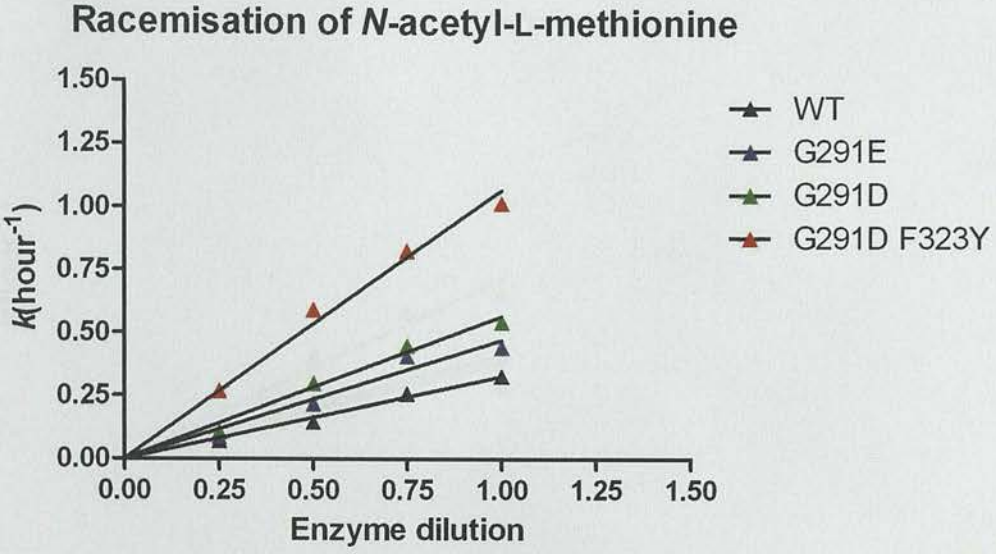


Figure 57. Plot of *N*-acetyl-L-methionine conversion ((transformed using *Eqn (1)*) versus NAAAR concentration.

Table 10. Rate constants (calculated with *Eqn (3)*) and specific activity (calculated with *Eqn (4)*) for whole cell biotransformations using different NAAAR variants.

Variant	*k (hour^{-1}) [R^2]*	*Specific activity (µmoles hour^{-1} OD^{-1})*
WT	0.1634 [0.98]	222 ± 14
G291E	0.2355 [0.95]	321± 11
G291D	0.2830 [0.98]	385 ± 14
G291D F323Y	0.5343 [0.99]	728 ± 19

***N*-acetyl-D-methionine:**

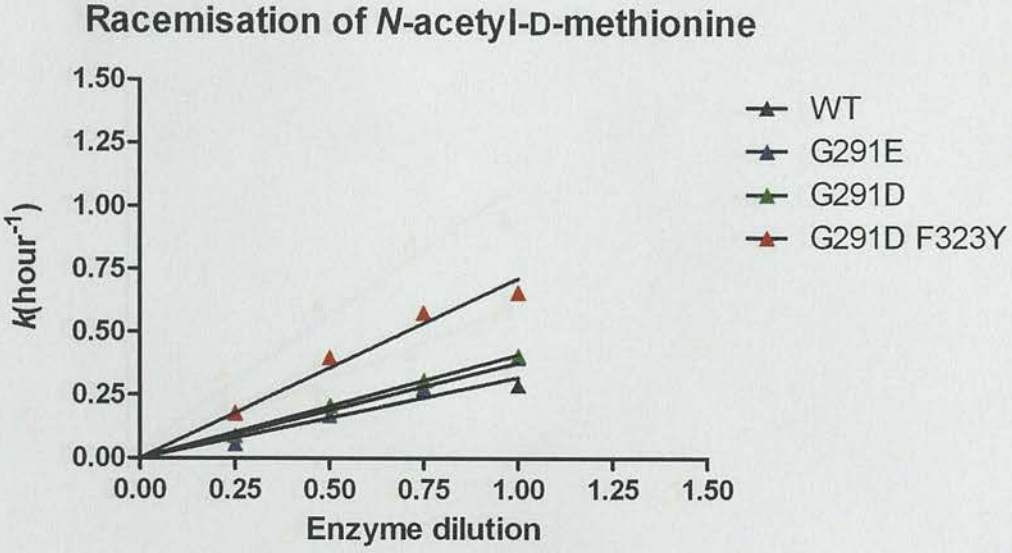


Figure 58. Plot of *N*-acetyl-D-methionine conversion ((transformed using *Eqn (2)*) versus NAAAR concentration.

Table 11. Rate constants (calculated with *Eqn (3)*) and specific activity (calculated with *Eqn (4)*) for whole cell biotransformations using different NAAAR variants.

Variant	*k (hour^{-1}) [R^2]*	*Specific activity (μmoles hour^{-1} OD^{-1})*
WT	0.1600 [0.99]	210 ± 7
G291E	0.1900 [0.94]	233 ± 4
G291D	0.2070 [0.97]	282 ± 7
G291D F323Y	0.3584 [0.96]	655 ± 12

***N*-acetyl-D-allylglycine:**

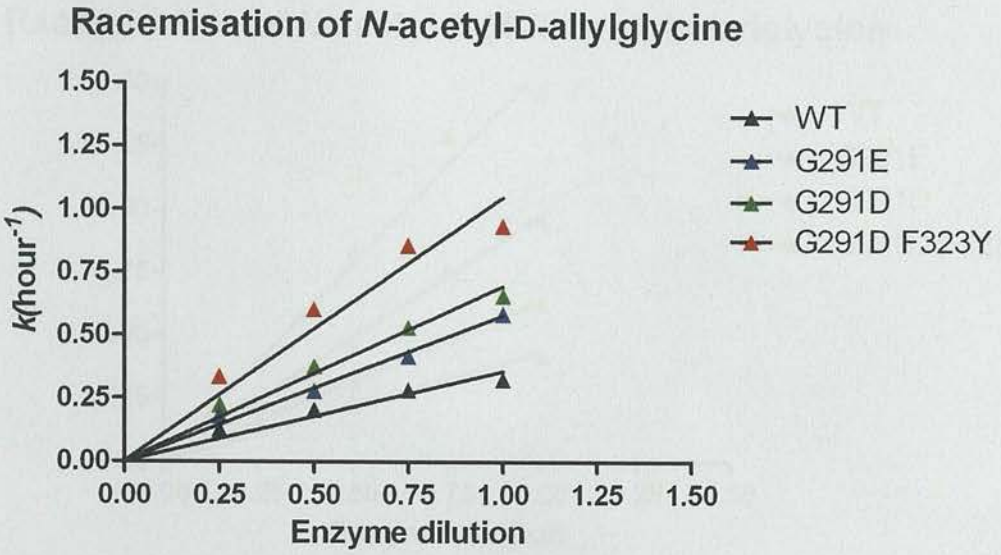


Figure 59. Plot of *N*-acetyl-D-allylglycine conversion ((transformed using *Eqn (2)*) versus NAAAR concentration.

Table 12. Rate constants (calculated with *Eqn (3)*) and specific activity (calculated with *Eqn (4)*) for whole cell biotransformations using different NAAAR variants.

Variant	*k (hour^{-1}) [R^2]*	*Specific activity (µmoles hour^{-1} OD^{-1})*
WT	0.1791 [0.85]	236 ± 9
G291E	0.2904 [0.98]	383 ± 14
G291D	0.3387 [0.95]	445 ± 9
G291D F323Y	0.5256 [0.87]	691 ± 14

***N*-acetyl-D-(4-fluoro)phenylglycine:**

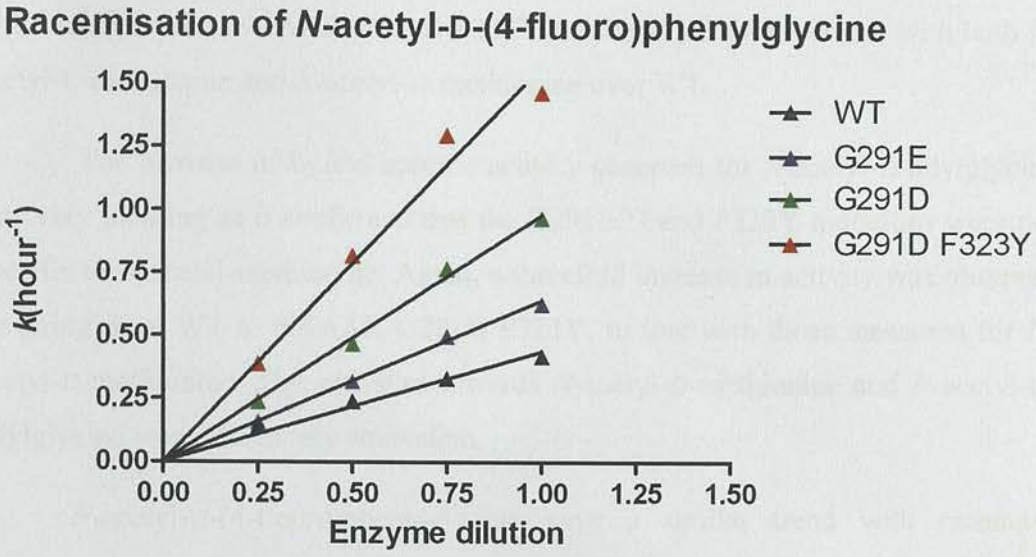


Figure 60. Plot of *N*-acetyl-D-(4-fluoro)phenylglycine conversion ((transformed using *Eqn (2)*) versus NAAAR concentration.

Table 13. Rate constants (calculated with *Eqn (3)*) and specific activity (calculated with *Eqn (4)*) for whole cell biotransformations using different NAAAR variants.

Variant	*k (hour^{-1}) [R^2]*	*Specific activity (μmoles hour^{-1} OD^{-1})*
WT	0.2185 [0.97]	287 ± 47
G291E	0.3218 [0.99]	395 ± 23
G291D	0.4914 [0.99]	670 ± 42
G291D F323Y	0.7878 [0.97]	1441 ± 60

The *in vivo* racemisation assays complemented the earlier observed *in vitro* results with an increase observed in the rate of racemisation (k_L and k_D) with each round of evolution. NAAAR G291D F323Y exhibited a 3 fold increase with both *N*-acetyl-L-methionine and *N*-acetyl-D-methionine over WT.

The increase in k_D and specific activity observed for *N*-acetyl-D-allylglycine was very pleasing as it confirmed that the G291E/D and F323Y mutations were not specific to *N*-acetyl-methionine. Again, a threefold increase in activity was observed on going from WT to NAAAR G291D F323Y, in line with those measured for *N*-acetyl-D-methionine. The activities towards *N*-acetyl-D-methionine and *N*-acetyl-D-allylglycine were effectively equivalent.

N-acetyl-D-(4-fluoro)phenylglycine gave a similar trend with racemase activity increasing with each improved NAAAR variant. However, in this case there was a 5 fold increase in specific activity. Activity towards this substrate was 120% higher than that towards *N*-acetyl-D-allylglycine and *N*-acetyl-D-methionine. Phenylglycines are also known to undergo chemical racemisation easily and this likely helps to increase the rate at which NAAAR can racemise these substrates. Chemical racemisation can occur under basic conditions due to the low pKa of the α-proton. However, the increases observed here were not thought to be occurring by non-enzymatic racemisation. All HPLC running buffers were acidic (~pH 4.5) to ensure that there was no pre- or on-column racemisation. During sample preparation, 5 M NaOH was used to solubilise hydrophobic amino acids, and this has the possibility of introducing a level of non-enzymatic racemisation. However, HPLC analysis of reactions containing no racemase, confirmed that racemisation (with all substrates) was dependent on the presence of NAAAR.

The L-enantiomers of both *N*-acetyl-allylglycine and *N*-acetyl-(4-fluoro)phenylglycine were not available for testing or for use as analytical standards for HPLC analysis. However, as the reaction generated a new peak on the chiral HPLC trace which became exactly 50% of the total HPLC area (with a decrease of 50% in the D-enantiomer peak), it is safe to assume that this new peak was the L-enantiomer – although not isolated and confirmed by NMR or MS.

4.4.3, Small scale dynamic kinetic resolutions (DKRs) of N-acetyl-DL-methionine using NAAAR/L-acylase whole cell biotransformations:

The specific activity of BL21(DE3) cells expressing NAAAR G291D F323Y has been found to be between three and five fold higher than those expressing the WT enzyme. However, the racemisation reaction must be coupled with a stereoselective acylase in the form of a DKR to be of industrial use. To test how the generated NAAAR variants compared to NAAAR WT in this reaction, small scale resolutions of *N*-acetyl-DL-methionine were carried out to yield enantiopure L-methionine. This was performed in a BL21(DE3) host expressing NAAAR and an L-acylase. All four NAAAR variants were expressed in separate BL21(DE3) hosts in LB broth (100 μg mL^{-1} ampicillin) over 24 hours with no IPTG for induction. L-acylase (Chirotech Technology Ltd) was expressed over 24 hours in auto-induction media (100 μg mL^{-1} ampicillin) with no IPTG from a proprietary plasmid[86]. Resolutions were initiated by adding equal amounts (0.2 OD_{600} units) of BL21(DE3) cells expressing either NAAAR (WT, G291E, G291D, or G291D F323Y) or L-acylase. The cells were mixed with 30 mM *N*-acetyl-DL-methionine, 5 mM $CoCl_2$, 50 mM Tris-HCl (pH 8.0) and incubated at 60°C. Samples were taken for chiral HPLC before addition of cells, and then every hour for five hours after addition. All four possible substrates/products of the reaction (*N*-acetyl-L-methionine, *N*-acetyl-D-methionine, L-methionine, and D-methionine) are well separated by chiral HPLC making it possible to measure the yield and e.e. of the L-methionine product.

The progress of each resolution is shown below; L-acylase only (**Figure 61**), NAAAR WT/L-acylase (**Figure 62**), NAAAR G291E/L-acylase (**Figure 63**), NAAAR G291D/L-acylase (**Figure 64**), and NAAAR G291D F323Y/L-acylase (**Figure 65**). As expected, with no NAAAR present, *N*-acetyl-D-methionine is not a substrate for the L-acylase. Inclusion of NAAAR WT resulted in a 20% conversion of *N*-acetyl-D-methionine to L-methionine. NAAAR G291E and NAAAR G291D increases this to 25% and 40% respectively, however, the L- and D- enantiomers were not being 'hydrolysed' at the same rate, *i.e.* $k_{hydrolysis} > k_{racemisation}$. When NAAAR G291D F323Y was added to the reaction conversion increased to 62% with both the conversion of *N*-acetyl-L-methionine and *N*-acetyl-D-methionine occurring at equivalent rates implying $k_{hydrolysis} \leq k_{racemisation}$. This would suggest that the reaction is running as an effective DKR[263]. No D-methionine was observed during chiral HPLC analysis indicating the e.e. of L-methionine in all reactions was >99%.

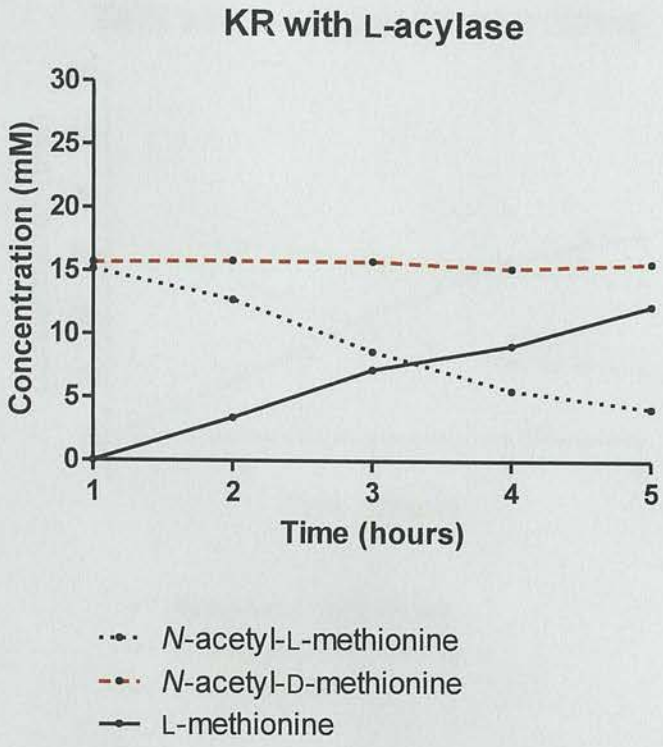


Figure 61. KR of *N*-acetyl-DL-methionine using L-acylase to yield L-methionine.

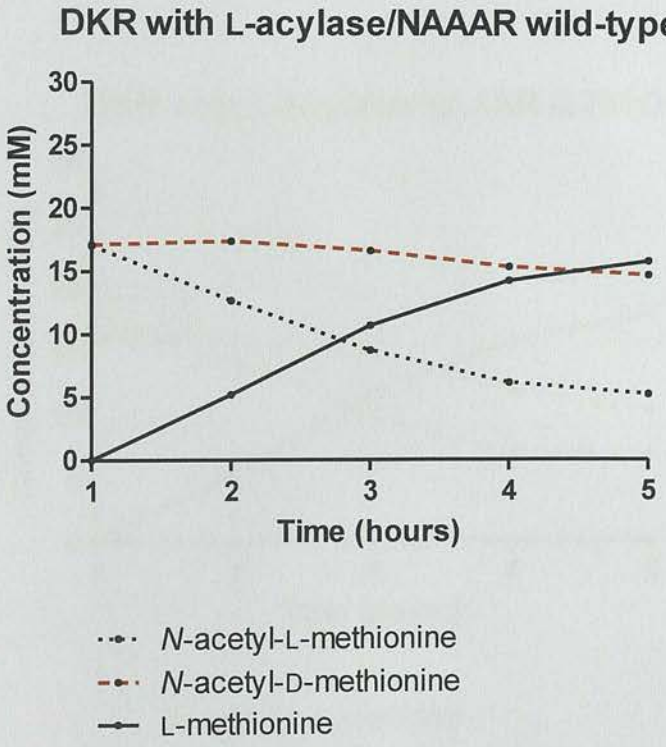


Figure 62. DKR of *N*-acetyl-DL-methionine using L-acylase/NAAAR WT to yield L-methionine.

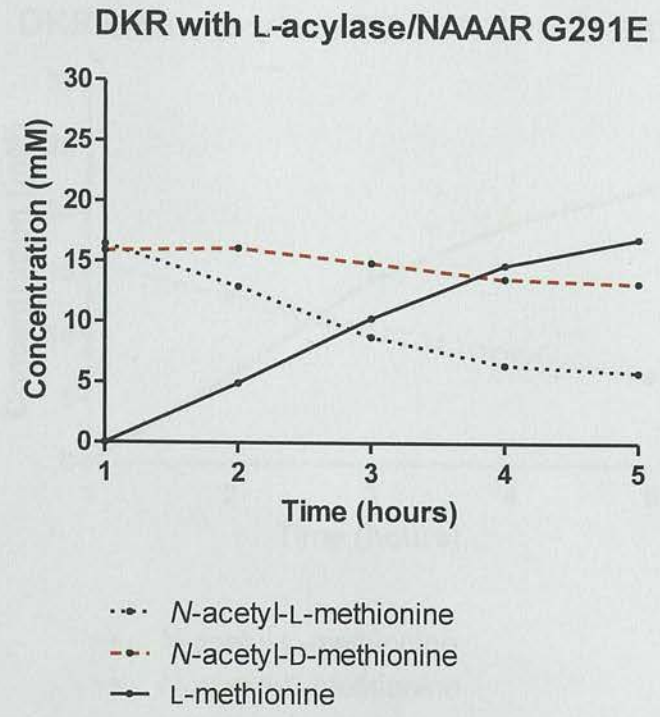


Figure 63. DKR of *N*-acetyl-DL-methionine using L-acylase/NAAAR G291E to yield L-methionine.

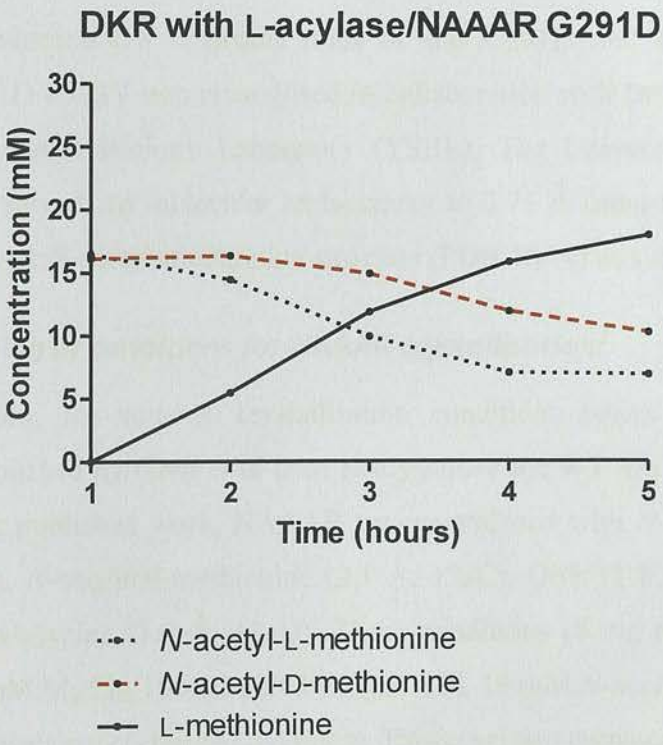


Figure 64. DKR of *N*-acetyl-DL-methionine using L-acylase/NAAAR G291D to yield L-methionine.

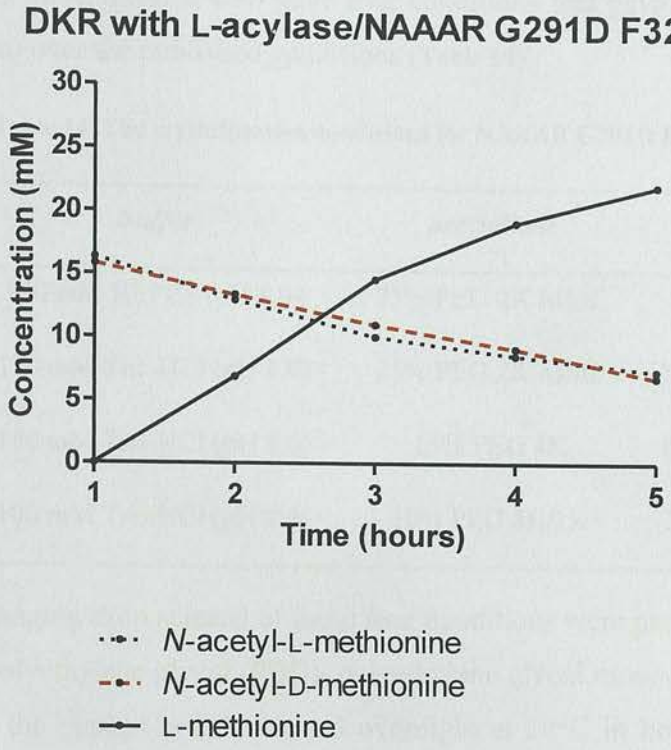


Figure 65. DKR of *N*-acetyl-DL-methionine using L-acylase/NAAAR G291D F323Y to yield L-methionine.

4.5, Structural biology of NAAAR G291D F323Y:

To determine the chemical roles of the G291D and F323Y mutations, NAAAR G291D F323Y was crystallised in collaboration with Dr Gideon Grogan at the York Structural Biology Laboratory (YSBL), The University of York. The structure was solved *via* molecular replacement to 2.71 Å using the published *Am. Ts-1-60* NAAAR:*N*-acetyl-methionine structure (PDB:1SJA) as a model[100].

4.5.1, Screening of conditions for NAAAR crystallisation:

Screening for suitable crystallisation conditions began with testing the conditions published by Gerlt *et al* used to crystallise the WT *Am. Ts-1-60* NAAAR protein. In the published work, NAAAR was crystallised with *N*-acetyl-methionine (2.3 Å, 1SJA), *N*-succinyl-methionine (2.1 Å, 1SJC), OSB (2.2 Å, 1SJB), and *N*-succinyl-phenylglycine (1.9 Å, 1SJD). These conditions (8 mg ml^{-1} NAAAR, 8% PEG 8K, 10 mM $MgCl_2$, 100 mM HEPES (pH 8.0), 10 mM *N*-acetyl-DL-methionine) gave large crystals which formed overnight. However, two commercial screens (CSS and Index) were also used to check for other conditions that may give better-quality

crystals. From these screens, CSS gave four conditions that gave improved crystals (form and size) over the published conditions (**Table 14**).

Table 14. The crystallisation conditions for NAAAR G291D F323Y.

Condition	*Buffer*	*precipitant*	*co-precipitant*
A1	100 mM HEPES (pH 8.0)	25% PEG 2K MME	300 mM Na acetate
A12	100 mM Tris-HCl (pH 8.0)	25% PEG 2K MME	800 mM Na formate
B12	100 mM Tris-HCl (pH 8.0)	15% PEG 4K	800 mM Na formate
C7	100 mM Tris-HCl (pH 8.0)	10% PEG 8K/1K	300 mM Na acetate

Large scale hanging drop screens of these four conditions were performed with a ± 1 % spread in polyethylene glycol (PEG)/ polyethylene glycol monomethyl ether (PEG MME) along the plate. Crystals formed overnight at 20°C in both apo- and holo-samples with no discernable differences between either. Samples from B12 gave the lowest diffraction (3.0 Å) with no cryo-protectant. Four crystals from these conditions were flash cooled in liquid nitrogen prior to data collection at the synchrotron.

4.5.2, Processing and refinement of NAAAR G291D F323Y data:

All structural processing and refinement was performed by Dr Gideon Grogan (YSBL, The University of York).

4.5.3, Structure of NAAAR G291D F323Y:

The crystal structure of NAAAR G291D F323Y (PDB code: 4A6G) revealed two new, intra-molecular interactions between the introduced side groups of D291 and Y323 with the native R299 (**Figure 66**). The electron density map of ligand (*N*-acetyl-L-methionine), Y323, D291, and R299 is shown below (**Figure 67**). Due to the lower resolution of the data compared to the published structure (2.7 Å compared to 2.3 Å), conclusions regarding any protein:ligand interactions which may occur in the variant structure are hard to draw. However the substrate did appear to be slightly displaced towards the Mg^{2+}. No significant differences were noted between the quaternary and tertiary structures of NAAAR G291D F323Y and 1SJA. The asymmetric unit contained four monomers, half of the octameric structure observed by size exclusion chromatography.

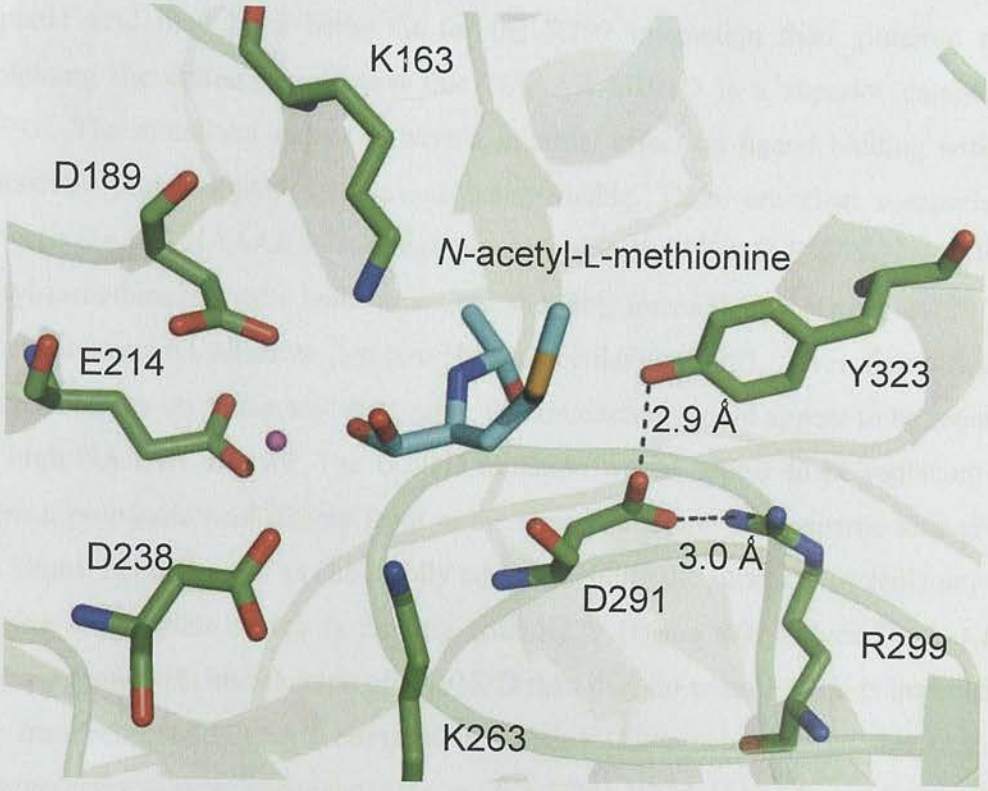


Figure 66. Active site of *Am. Ts-1-60* NAAAR G291D F323Y showing new salt bridge interaction between D291 and R229 and H-bond interaction between Y323 and D291. Mg^{2+} ion shown as magenta sphere.

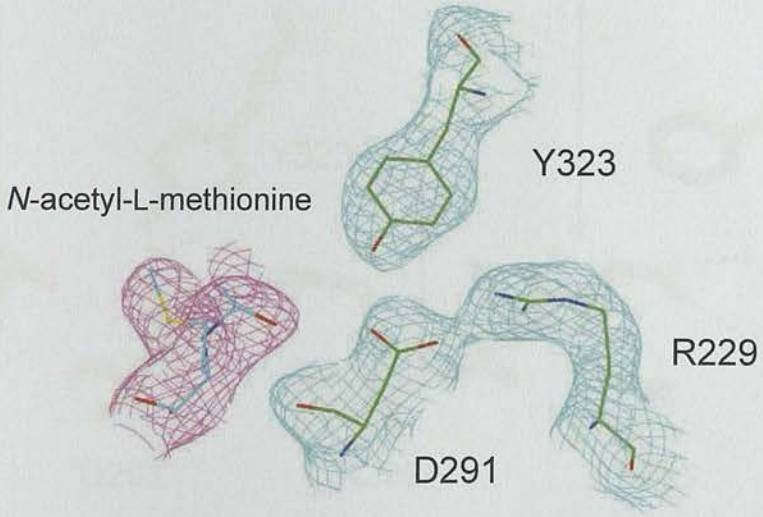


Figure 67. Electron density (2.7 Å) in cyan, for of D291, R299 and Y323 of NAAAR D291 Y323 and ligand electron density shown in pink, corresponding to the omit map (*Fo-Fc* map).

The major impact of the G291D and F323Y mutations has been the formation of two new intra-molecular interactions. These are a 3.0 Å salt bridge between the carboxylate side group of D291 and the guanidinium of R299, along with a 2.9 Å H-bonding interaction between the hydroxyl of F323 and the carboxylate of D291.

Aspartic acid may be a better fit for the R299 interaction than glutamic acid, explaining the earlier observation that NAAAR G291D is a superior catalyst to G291E. The mutations appear to have a minimal effect on ligand binding with no immediate protein:ligand interactions being visible. There are clear comparisons between the WT NAAAR:*N*-succinyl-D-methionine and G291D F323Y NAAAR:*N*-acetyl-L-methionine with both structures showing interactions between R299 and either the succinyl substrate (*via* two H_2O molecules) or D291. Given that NAAAR is most activity on *N*-succinyl substrates, this interaction would appear to be required for high NAAAR activity. The G291D mutation would appear to be replacing the missing propanoic acid moiety from in the acetyl structure. The aspartic side group can almost be thought of as chemically complementing the substrate by replacing the missing carboxylate which is binding with R229 (**Figure 68**). Values for k_{cat} also increase more with introduction of G291E/D than they do when F323Y is introduced. The occurrence of NAAAR enzymes in which F323 is naturally tyrosine (*e.g. G. kaustophilus*) is now understandable as, the D291-Y323 H-bond could also form Y323 to the succinyl substrate carboxylate.

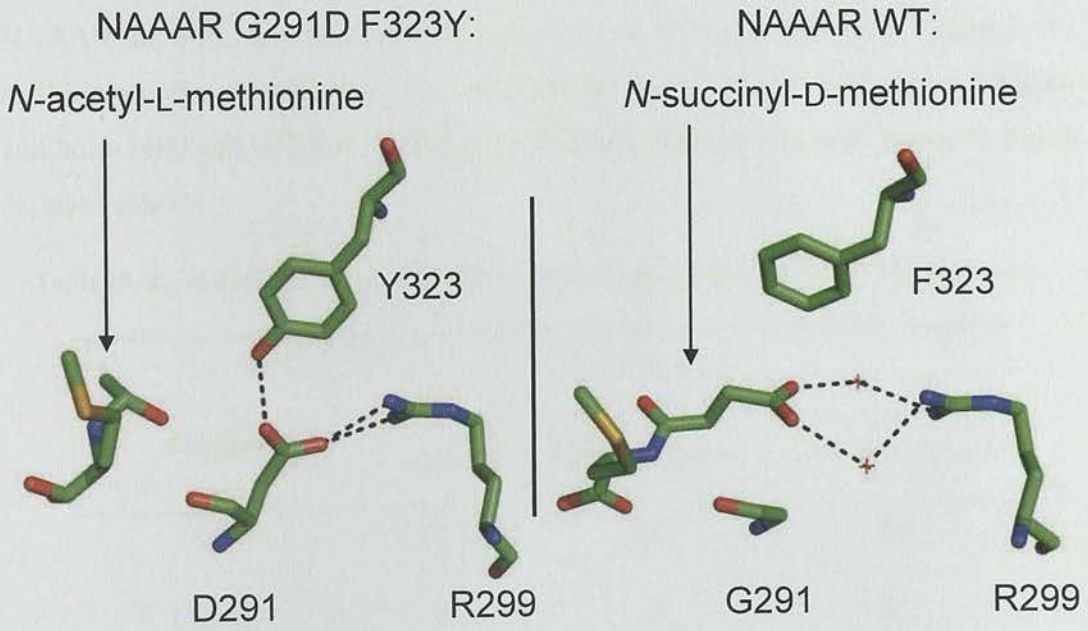


Figure 68. Comparison of NAAAR G291D F323Y: *N*-acetyl-L-methionine and WT NAAAR: *N*-succinyl-D-methionine. H_2O molecules shown as red crosses in WT NAAAR:*N*-succinyl-D-methionine structure (PBD:1SJC)[100]. All bonding interactions ~3 Å.

The introduction of salt bridges, like the D291-R299 interaction, has been shown to have an effect on catalytic performance[264]. To determine the effect of these mutations on protein stability, the melting temperatures (T_m) of each NAAAR variant and the WT was determined.

4.5.4, Melting temperatures (T_m) of WT and variant NAAARs:

The T_m of a protein is point at which half the protein molecules are folded and half un-folded[265]. The temperature at which this occurs is taken as a measure of how thermostable that protein is. T_m is determined by measuring the florescence signal of a dye (SYPRO orange) as a function of temperature[266]. When the protein is fully folded and in its native state, the fluorescence signal is quenched by the aqueous environment, but as temperature increases and the protein unfolds, the internal hydrophobic patches are revealed. These bind the dye and increase SYPRO orange fluorescence. T_m was determined for each NAAAR variant in both the apo- and holo- state to measure the effects of the new interactions introduced into the NAAAR active site. This was performed on a BioRad iQ5 rtPCR Thermocycler at the Edinburgh Protein Production Facility. T_m was calculated with 0.5 mg mL^{-1} NAAAR in both the apo- and holo-form with 100 and 200 mM *N*-acetyl-DL-methionine. The calculated T_m for each enzyme and T_m differential curves for apo- and holo- (100 and 200 mM ligand) shown is shown below (**Figure 69, Figure 70, Figure 71,** and **Table 15**).

Table 15. T_m of each NAAAR in the apo- and holo-form. Ligand = *N*-acetyl-DL-methionine.

Variant	*T_m (°C)*		
	Apo	*100 mM ligand*	*200 mM ligand*
WT	67	68	69
G291E	70	71	72
G291D	71	72	73
G291D F323Y	72	75	76

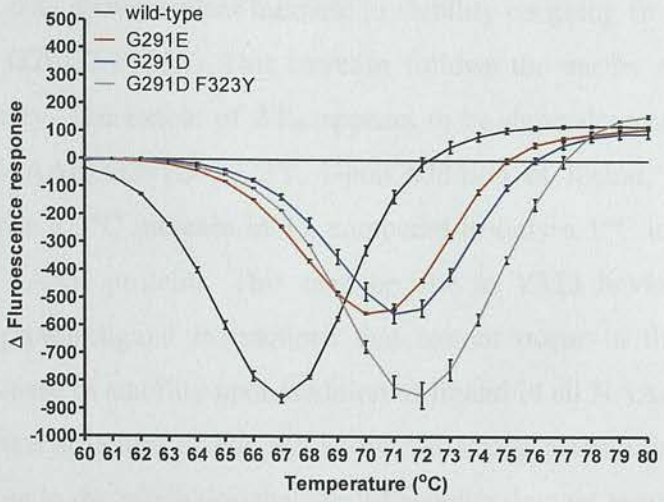


Figure 69. Differential graph of fluroescence response against temperature for apo-NAAAR.

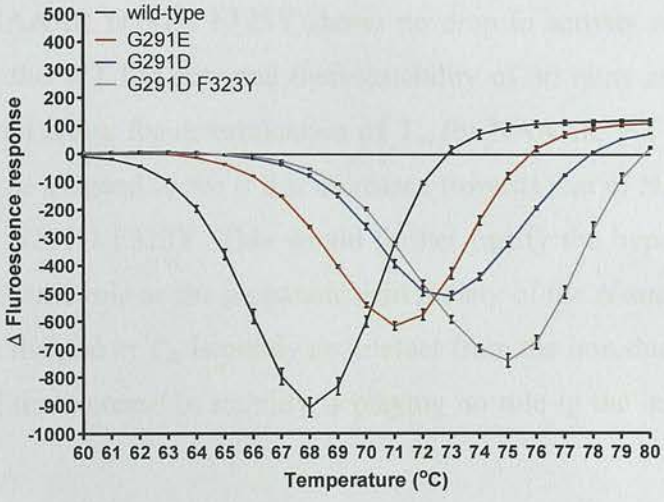
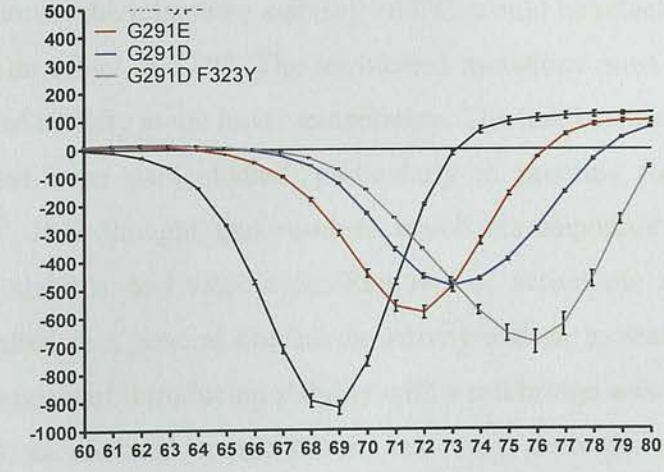


Figure 70. Differential graph of fluroescence response against temperature for holo-NAAAR (100 mM *N*-acetyl-DL-methionine).



The T_m data shows a clear increase in stability on going from WT → G291E → G291D → G291D F323Y. This increase follows the earlier observed trend of increasing activity. The extent of ΔT_m appears to be dependent on the presence of substrate in NAAAR G291D F323Y. Upon addition of ligand, NAAAR G291D F323Y undergoes a 3°C increase in T_m compared to only a 1°C in the WT, G291E and G291D NAAAR proteins. This may be due to Y323 having introduced the possibility of protein:ligand interactions that cannot occur in the F323 proteins. There is an increase in stability upon addition of ligand in all NAAAR variants, even though the crystal structure of the WT shows no strong protein:ligand interactions. This is likely due to the prediction that the lid/capping domain remains un-ordered in the absence of ligand, and hence is less stable[98-100]. The increase in T_m would also explain why NAAAR G291D F323Y shows no drop in activity after five hours at 60°C, whereas the WT has reported thermostability of 30 mins at 55°C. Time and resources did not allow for determination of T_m for NAAAR WT with *N*-succinyl-DL-methionine as a ligand to see if it is increased towards that of NAAAR G291D or even NAAAR G291D F323Y. This would further justify the hypothesis that D291 performing the same role as the propanoic acid moiety of the *N*-succinyl substrate. It is possible that the rise in T_m is purely an artefact from the introduction of this R299 interaction and the increase in stability is playing no role in the increase in enzyme efficiency.

It should be kept in mind that these mutations have been selected for during a genetic selection process performed at 37°C. There appears to be no reason why a series of mutations, which increase stability >69°C, would be selected for unless they can also exert an effect at 37°C. The introduced mutations must also be having a positive effect of activity at the lower temperature. The link between protein stability and activity has been well studied, particularly in proteins from psychrophilic bacteria[267, 268]. It is thought that residues which are important for function are detrimental to stability and *vice versa*. Removal of active site residues from T4 Lysozyme resulted in a general decease in activity and an increase in stability[269]. The opposite process of introducing stability with a salt bridge was shown to increase thermostability at elevated temperatures, but then decrease activity at lower temperatures, of an Acylphosphatase (PhAcP, EC 3.6.1.7) from *Pyrococcus horikoshii*[264]. The literature would suggest that in general, an increase in stability of an active site will lower its activity at lower temperature (37°C). Clearly this is not

the case here as all three mutations show an increase in activity at both 37°C (*in vivo* screen) and 60°C (*in vitro* and *in vivo* assays), along with an increase in stability above 69°C (T_m measurements). These would appear to be in contrast with the general trend observed considering mutations that affect stability and activity.

4.6, Conclusions:

Using a genetic selection system directed evolution techniques have been successfully used to increase the activity of NAAAR from *Am. Ts-1-60*. Improvements in the genetic selection method, such as a change in minimal media and expression plasmid removed a sizable variation in colony size that had been hindering selection and generating a large number of false positives. Although activity was evolved against *N*-acetyl-D-methionine, an increase in activity was observed towards both enantiomers of this substrate, along with several other natural and non-natural *N*-acetyl-amino acids.

A double NAAAR variant containing G291D and F323Y was found to have a 4.7 fold increase in k_{cat} for *N*-acetyl-DL-methionine, although no major change in K_m was observed. Saturation mutagenesis at these positions highlighted that they were the most effective mutations. Analysis of the NAAAR WT crystal structure revealed that both mutations are found within the acyl binding pocket, in close proximity to the substrate. To decipher the molecular basis of the increase in activity, the NAAAR G291D F323Y variant was crystallised. The structure was solved to 2.71 Å with molecular replacement using the published WT structure. The variant structure showed two new intramolecular interactions that introduced a 3.0 Å salt bridge between D291 and R229 and a 2.9 Å H-bonding interaction between Y323 and D291. These interactions suggested that the enzyme has been stabilised and that this may be source of increased catalytic activity. Literature evidence suggested that this could be the case given the lack of obvious protein:ligand interactions. The change in protein stability was measured by determining the T_m for the WT and variant NAAARs. A 7°C increase in T_m was measured for NAAAR G291D F323Y compared to the WT (69°C) in the presence of 200 mM ligand (*N*-acetyl-DL-methionine). As the increase in T_m between the WT and the G291D F323Y variant was greater in the presence of substrate (only 5°C in the apo- form), this would suggest that ligand:D291/Y323 interactions have also been introduced, although

these are not immediately clear in the solved structure. This may be due to the relatively low resolution of the crystals (2.71 Å).

4.7, Future work:

The 4.7 fold increase in activity shown by the G291D and F323Y mutations could likely be built upon by further mutagenesis. Whether this should target protein:ligand interactions or look to further increase the protein's stability is unclear. An obvious target for mutagenesis is M50 and its surrounding neighbours which appear to be responsible for amino acid substrate specificity. Mutagenesis of these residues will require a different screening/selection strategy than the SET21 genetic selection employed here which is specific for *N*-acetyl-D-methionine. High throughput chiral HPLC screening was successfully performed to carry out saturation mutagenesis at F323 confirming that F323Y was the most active residue – applying this approach these residues would be relatively simple to implement with the correct equipment and assays. As NAAAR has a very broad substrate range, accepting most amino acids with varying levels of activity, the substrate R-group binding pocket could likely be optimised to suit a specific amino acid if a mutagenesis program was focused on this area of the protein. Changes in the *N*-acyl pocket could be easily accessed by synthesis of the appropriate methionine derivatives, *i.e.* *N*-benzoyl-D-methionine for evolution of an *N*-benzoyl amino acid racemase.

Although not characterised fully in this project, two further NAAAR enzymes for *Geobacillus* have also been cloned from *G. kaustophilus* and *G. thermodenitrificans*. These enzymes may provide a more active NAAAR starting point for mutagenesis, or have different substrate preferences to the *Am. Ts-1-60* NAAAR used here. Interestingly, both *Geobacillus* NAAARs are Y323, which may be forming H-bonds between the succinyl carboxylate and Y323 like the D291-Y332 H-bond observed with NAAAR G291D F323Y:*N*-acetyl-L-methionine. To determine whether *N*-succinyl racemase activity has been affected, and to support the hypothesis regarding the molecular role of these mutations, the activity of NAAAR containing the F323Y and G291D mutations should be measured with an *N*-succinyl amino acid. Given that *Geobacillus* NAAARs already contain F323Y, it may be that this mutation increases activity with *N*-succinyl substrates as it can interact with the succinyl carboxylate as it does with the aspartic carboxylate. If the D291 side group

is acting as a mimic for the succinyl carboxylate, variants with this mutation may no longer accept the *N*-succinyl substrate.

Chapter 5, Industrial applications of Amycolatopsis Ts-1-60 NAAAR G291D F323Y in the resolution of enantio-pure amino acids:

5.1, Industrial application of Amycolatopsis NAAAR G291D F323Y:

The goal of this project was to generate a cost effective racemase biocatalyst that would fit in to the established D/L-acylase process for production of enantiopure α-amino acids. Given the promise shown by the coupled NAAAR G291D F323Y/L-acylase whole cell biocatalyst DKR of *N*-acetyl-DL-methionine (See section 3.3.3), and the marked increase in activity of NAAAR G291D F323Y compared to the WT, there was strong reason to think that NAAAR G291D F323Y would meet these criteria. To study the performance of NAAAR G291D F323Y under industrial conditions, the enzyme was first prepared by fermentation, and the resulting biocatalyst tested under various reaction conditions including NAAAR/D-acylase DKRs of several *N*-acetylated amino acids.

5.1.1, Fermentation of Amycolatopsis NAAAR G291D F323Y:

Biocatalysts are routinely produced by fermentation to allow for high cell densities and large volumes of enzyme. *E. coli* grown in a shake flask will reach an OD_{600} of the range 5-10, whereas a fermentation of the same host will reach 100-150 in the same time scale. Some exceptional fermentation expression hosts, such as *Pseudomonas fluorescens,* marketed as the *Pfenex* expression system, will generate an OD_{600} of >400.

A 1 L fermentation of BL21(DE3) expressing NAAAR G291D F323Y was carried out over 48 hours at 30°C. Similar to shake flask expression for protein purification no IPTG was added (See section 3.3.2.1). The fermentation process started with a 50 mL overnight culture of BL21(DE3) pET20b NAAAR G291D F323Y in *E. coli* seed media. This was then used to inoculate 700 mL of *E. coli* fermentation media. After 19.5 hours, *E. coli* feed media was dripped in at a rate of 14 mL $hour^{-1}$. Expression of NAAAR was monitored by measuring NAAAR activity and SDS-PAGE analysis (**Figure 72**). After 48 hours, the fermentation was harvested with centrifugation (*5,000 g*, 30 mins, 4°C), giving a yield of 92 g of wet cells (16.2 g dry cells) from 1 L of media.

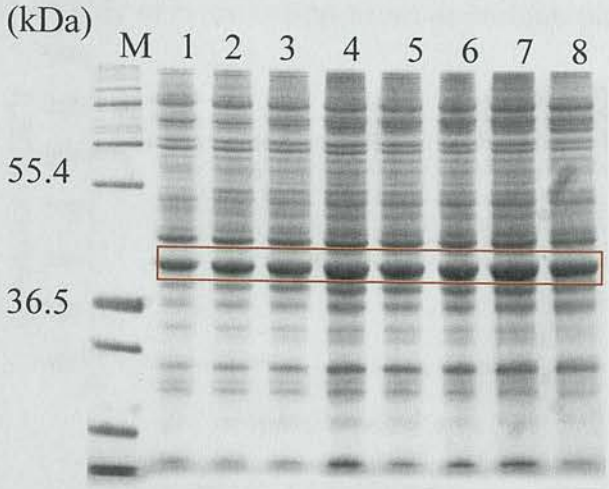


Figure 72. SDS-PAGE analysis of BL21(DE3) pET20b NAAAR G291D F323Y fermentation. NAAAR (Mr = 39.4 kDa) expression highlighted in red box. M= Marker, 1= fermentation after 17 hours, 2 = fermentation after 18.5 hours, 3 = fermentation after 20 hours, 4 = fermentation after 21 hours, 5 = fermentation after 22 hours, 6 = fermentation after 23 hours, 7 = fermentation after 24 hours, 8 = fermentation after 40 hours.

The total NAAAR activity was calculated by means of the standard NAAAR activity assay. This involved incubation of a 10 fold dilution (with 50 mM Tris-HCl (pH 8.0)) of fermentation broth with 150 mM *N*-acetyl-L-methionine in 50 mM Tris-HCl (pH 8.0), 5 mM $CoCl_2$. Again, the substrate was dissolved in 100 mM Tris-HCl (pH 8.0) to control pH. The reaction was run at 60°C with agitation for 60 mins before termination and 1000 fold dilution with HPLC running buffer (75:25 0.01% TEAA:MeOH). Chiral HPLC was used to measure racemase activity, which was then converted in to a specific activity per mL of broth. One unit (U) of NAAAR activity was defined as the amount of enzyme required to racemise 1 µmole of *N*-acetyl-L-methionine in 60 minutes at 60°C. The specific and total activity were calculated for samples taken throughout the fermentation process indicating that specific activity (per OD_{600}) reached a maximum after 24 hours, with total activity reaching a maximum after 40 hours (**Figure 73** and **Figure 74**). The fermentation was harvested after 40 hours at an OD_{600} of 96. The total NAAAR activity present in the 92 g of cells was approx. 10,000,000 units. Specific NAAAR activity dropped after 24 hours, this may be due to the loss of antibiotic selection (100 µg mL^{-1} ampicillin). Addressing this would yield an improved NAAAR yield from the fermentation.

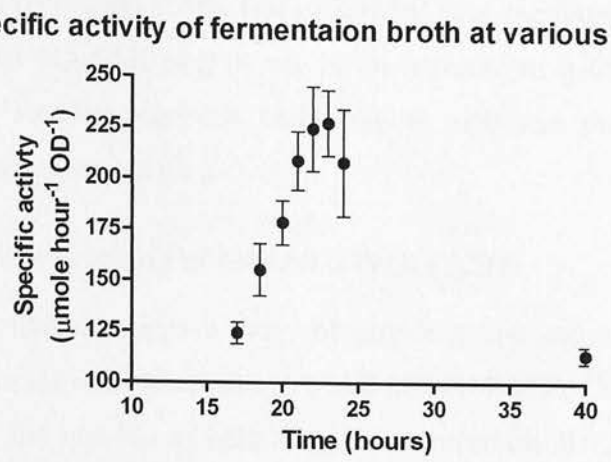


Figure 73. Specific NAAAR activity of fermentation broth at differing time points. Note that scale starts at 100 µmole hour^{-1} OD^{-1}.

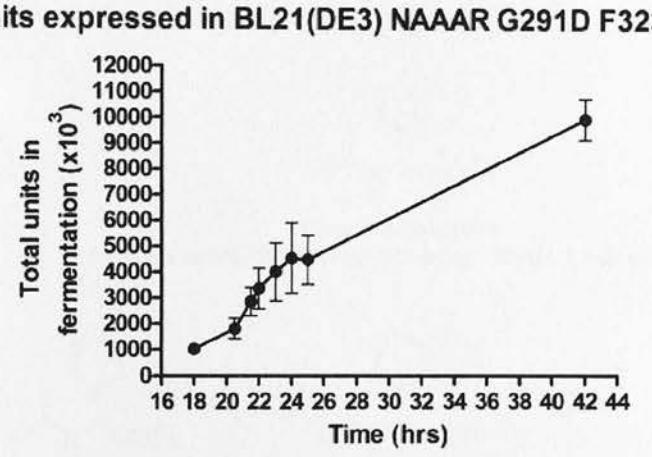


Figure 74. Total NAAAR activity present in BL21(DE3) pET20b NAAAR G291D F323Y fermentation broth.

5.1.2, Downstream processing (DSP) of Amycolatopsis NAAAR G291D F323Y biocatalyst:

BL21(DE3) pET20b NAAAR G291D F323Y fermentation broth was processed into a CFE from wet cells to allow it to be tested for industrial use. This was carried out by sonication (15 s on/ 15 s off) for 15 minutes at 2°C and then clarified by centrifugation (*12,000 g,* 30 mins, 4°C). The extract was tested for NAAAR activity using the standard NAAAR activity assay. 50 mL of CFE was generated from 10 g of wet cell paste, which had a specific activity 6615 (± 660) U mL^{-1}, where 1 U = the amount of enzyme required to racemise 1 µmole of *N*-acetyl-L-methionine in 1 hour at 60°C. DSP of all 92 g of wet cells would generate 460 mL of CFE with a total activity of 3,043,000 (± 304,000) U. This DSP was non-

optimised and represents a 70% loss of activity from the fermentation broth. Further mutagenesis of NAAAR may result in an increase in activity; however a more obvious and simpler approach would be to optimise the NAAAR DSP and fermentation to recover more activity.

5.1.3, Substrate profiling of NAAAR G291D F323Y:

The activity towards a range of proteinogenic and synthetically useful *N*-acetyl-amino acids was tested with NAAAR G291D F323Y CFE. Activity was tested by measuring the number of units required to racemise 300 mM (30 µmoles) of a single enantiomer of various substrates in 8 hours. These reactions were carried out at 60°C with agitation in 50 mM Tris-HCl (pH 8.0) with 5 mM $CoCl_2$, and again the substrate was dissolved in 100 mM Tris-HCl (pH 8.0). The required NAAAR loading and final % e.e. are listed below (**Figure 75**).

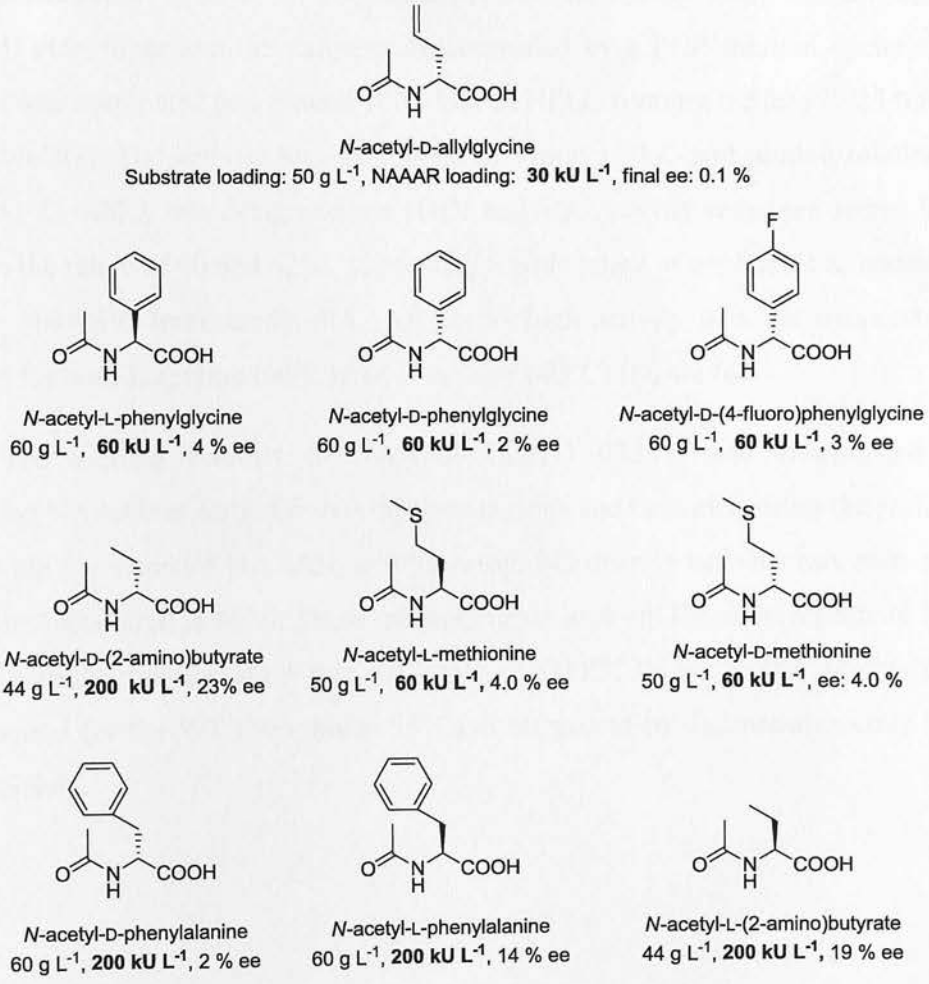


Figure 75. The required substrate loading for racemisation/partial racemisation of various substrates at 300 mM concentration.

The most active substrates were found to be three non-natural amino acids, *N*-acetyl-allylglycine, *N*-acetyl-phenylglycine, and *N*-acetyl-(4-fluoro)phenylglycine followed by *N*-acetyl-methionine. *N*-acetyl-D-allylglycine proved to be the best substrate with only 30 kU L^{-1} of NAAAR required to take the substrate to <1% e.e. *N*-acetyl-phenylalanine and *N*-acetyl-(2-amino)butyric acid proved to be poor substrates that did not form a racemate with 200 kU L^{-1} NAAAR. NAAAR activity towards *N*-benzoyl-DL-phenylalanine was also tested but no activity was measured with either enantiomer. No activity was detected towards *N*-acetyl-D-naphthylalanine or *N*-acetyl-L-orthinine.

5.1.4, Temperature studies of NAAAR G291D F323Y:

WT *Am. Ts-1-60* NAAAR is known to be a thermostable protein with reported activity up to 55°C[95]. This was further investigated by activity measurements from 35 to 80°C. The standard NAAAR activity assay was carried out in 50 µL PCR tubes and the temperature controlled by a PCR thermal cycler. The reaction was terminated and diluted 1000 fold in HPLC running buffer (75:25 0.01% TEAA:MeOH). The activity was calculated by chiral HPLC and plotted relative to that at 51°C, which was designated as 100% activity. Activity remained above 90% between the range of 40 and 62°C, suggesting a wide range of applicable temperature uses for NAAAR. Importantly, NAAAR shows high activity with the temperatures required for both L-acylase (60°C) and D-acylase (40°C) (**Figure 76**).

The thermal stability of NAAAR G291D F323Y was investigated by incubating NAAAR at 60°C for varying time periods and then measuring the residual activity *via* the standard NAAAR activity assay. No drop in activity was seen after five hours incubation at 60°C. These measurements back-up literature reports of high thermal stability, but also show that NAAAR G291D F323Y has higher stability than that reported for the WT (30 mins at 55°C) as suggested by T_m measurements (See section 3.4.4).

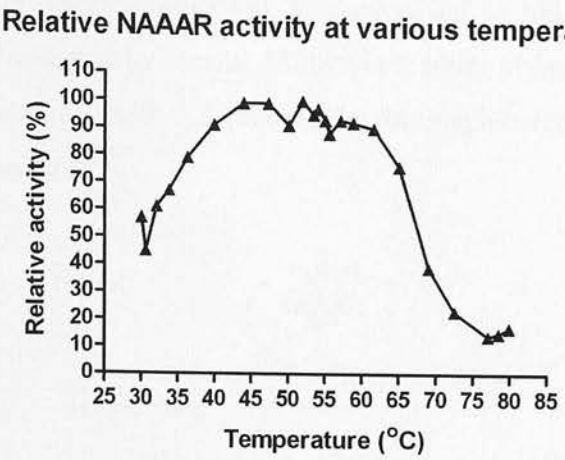


Figure 76. Relative NAAAR activity between 30°C and 80°C with *N*-acetyl-L-methionine.

5.2, NAAAR/D-acylase DKR processes:

NAAAR G291D F323Y was coupled to a D-acylase (Chirotech Technology Ltd) to measure the possible yields of D-amino acids using these enzymes. Both these enzymes require only $CoCl_2$ as a cofactor. A whole cell DKR process using NAAAR G291D F323Y/L-acylase had already been shown to work efficiently over 5 hours with 30 mM of *N*-acetyl-DL-methionine, however, this substrate concentration is too low to make this an economical process. To address this, industrially relevant substrate concentrations of 300 mM were used. This equates to roughly 50 g L^{-1} for an average *N*-acetylated amino acid. NAAAR/D-acylase DKRs were performed on *N*-acetyl-DL-serine, *N*-acetyl-DL-allylglycine, *N*-acetyl-DL-methionine, and *N*-acetyl-DL-phenylglycine (**Figure** 77).

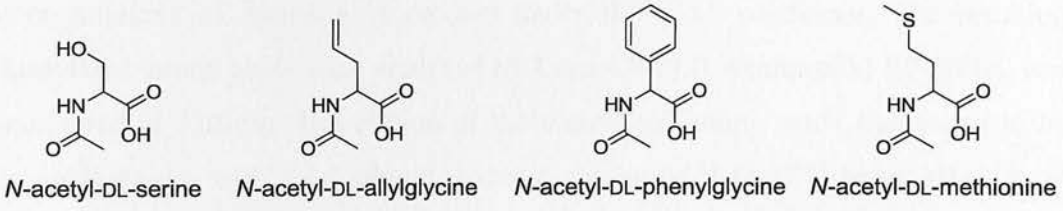


Figure 77. *N*-acetyl substrates for NAAAR/D-acylase coupled DKRs.

The D-acylase process represents a simple method to generate D-amino acids which cannot be fermented in the same manner as their L-enantiomers. The efficiency of NAAAR G291D F323Y would be judged by its ability to improve this reaction above the inherent 50% limitation. To mimic an industrial process, 300 mM substrate was used with the fermentation CFE, and reactions ran for 24 hours at 40°C. As D-serine and D-allylglycine lack a natural chromophore for UV detection, it

was necessary for these products to be derivatised to aid their detection. These reactions were terminated by careful dilution in a series of buffers to facilitate a final amine derivatisation with 5-(dimethylamino)naphthalene-1-sulfonyl chloride (dansyl-Cl) (**Figure 78**)[270].

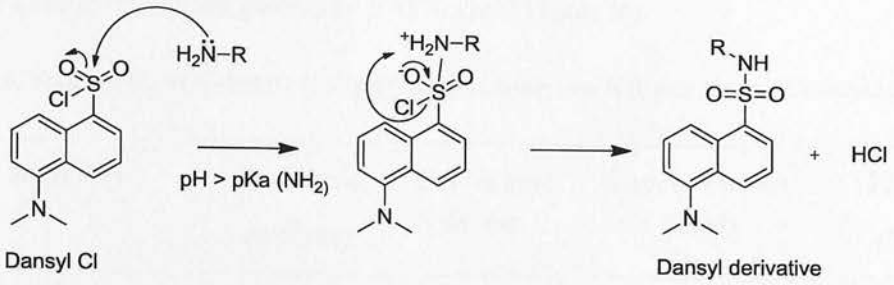


Figure 78. Dansylation of primary amines.

To allow the dansylation to be performed, the biotransformation was first terminated and diluted tenfold in H_2O:ACN (75:25) containing 10 mM DL-phenylalanine as an internal standard (I.S.) to monitor the efficiency of the dansylation process. The I.S. was used to generate the correction factor. Phenylalanine was selected as the internal standard as it was likely to have a different retention time to the amino acids being analysed. This tenfold diluted sample was then diluted tenfold again in H_2O:ACN (75:25), before a final tenfold dilution with 100 mM sodium carbonate buffer (pH 9.8). Dansyl derivatisation was performed by addition of 100 µL dansyl-Cl (3 mg mL^{-1}) in 100% acetone solution to the carbonate buffered biotransformation. The reaction was then incubated at 37°C for three hours before termination with 100 µL of glacial acetic acid. A standard curve was generated by derivatisation of known concentrations of amino acid product under the same conditions. The resulting dansylated amino acids were analysed by Luna C8(2) (Phenomenex) RP-HPLC and monitored at 330 nm. The elution of the dansylated amino acids was found to be straightforward, with a 15 minute isocratic elution of H_2O:ACN being effective at eluting all reaction species with excellent separation. Reactions were carried out on a 1 mL scale with varying amounts of NAAAR G291D F323Y CFE, typically between 30 kU L^{-1} and 120 kU L^{-1}, in 50 mM Tris-HCl with 5 mM $CoCl_2$. As always, substrate was prepared in 100 mM Tris-HCl (pH 8.0), giving a final Tris-HCl concentration of 100 mM.

5.2.4.1, NAAAR/D-acylase DKR of *N*-acetyl-DL-allylglycine:

NAAAR G291D F323Y coupled with D-acylase was found to be very efficient at producing D-allylglycine with only 30 kU L^{-1} of NAAAR required to take the reaction to completion (>99% D-allylglycine) in 24 hours. An equivalent D-acylase kinetic resolution gave only a 48% yield (**Table 16**).

Table 16. Final yields of *N*-dansyl-D-allylgylycine in D-acylase KR and NAAAR/D-acylase DKR.

Reaction	*Peak area (mVolts)*	*Correction factor*	*Concentration*[#] *(mM)*	*Yield (%)*
40 kU L^{-1} D-acylase	33.2	0.97	145.25	48.4%
40 kU L^{-1} D-acylase + 30 kU L^{-1} NAAAR G291D F323Y	72.1	0.92	298.65	99.5%

[#] Concentration = (peak area * correction factor) / 0.2218

RP-HPLC analysis of *N*-dansyl-D-allylglycine was performed with an isocratic elution (75:25 H_2O:ACN (0.01%TFA)). Un-reacted dansyl-Cl and the hydrolysis product dansyl-OH, eluted at 1 min. *N*-dansyl-D-allylglycine eluted at 4.15 mins and of *N*-dansyl-DL-phenylalanine eluted at 6.7 mins (**Figure 79**). A standard curve of *N*-dansyl-D-allylglycine was generated by dansylation of L-allylglycine standards (**Figure 80**).

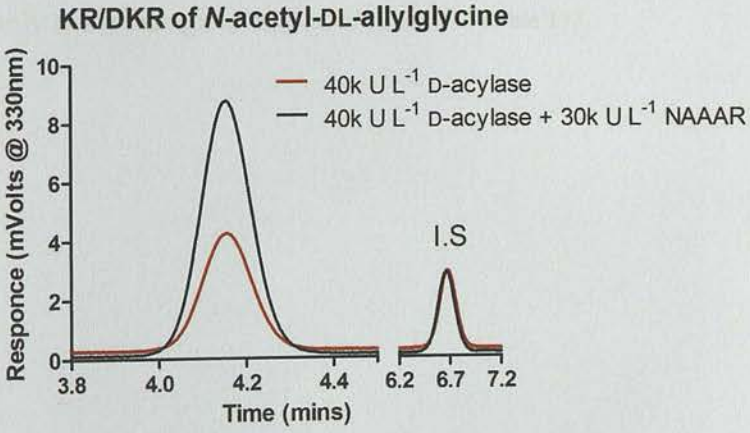


Figure 79. HPLC analysis (column = C8 Luna) of *N*-dansyl-D-allylglycine product from both D-acylase KR and NAAAR/D-acylase DKR.

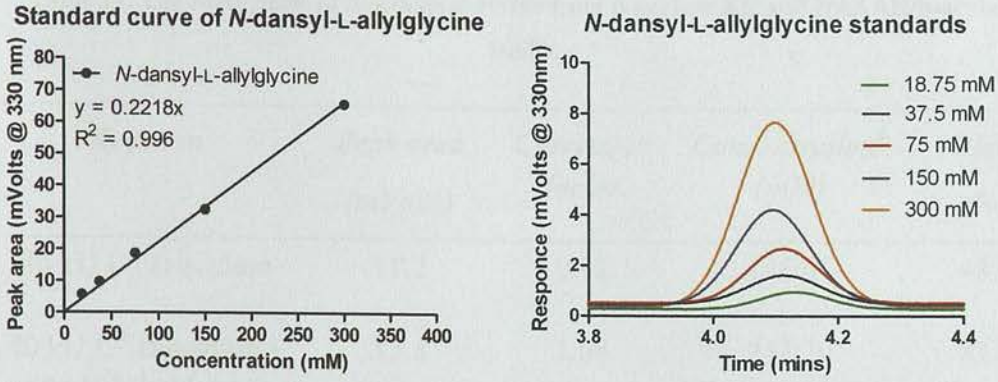


Figure 80. A Standard curve of *N*-dansyl-L-allylglycine (after 1000 fold dilution) response at 330 nm, and HPLC traces of standards (column = C8 Luna).

If all 3 million NAAAR units generated from DSP of a 1 litre fermentation were committed to a NAAAR/D-acylase resolution of *N*-acetyl-DL-allylglycine (100 L at 300 mM), 4.9 kg of *N*-acetyl-DL-allylglycine could be resolved to give a theoretical 3.6 kg of D-allylglycine in a 24 hour time frame. Of this, 1.8 kg would come from racemisation of *N*-acetyl-L-allylglycine before hydrolysis. This would represent an excellent economical yield with acetic acid as the only major by-product.

5.2.4.2, NAAAR/D-acylase DKR of *N*-acetyl-DL-serine:

Using the same reaction conditions (40 kU L^{-1} D-acylase, 40°C, 24 hours), 60 kU L^{-1} of NAAAR G291D F323Y was found to give an 81% yield of D-serine from 300 mM *N*-acetyl-DL-serine. Using 120 kU L^{-1} resulted in 99% yield, and a kinetic resolution with only D-acylase gave only a 49% yield (**Table 17**).

Table 17. The final yields of *N*-dansyl-D-serine from D-acylase KR and NAAAR/D-acylase DKRs.

Reaction	*Peak area (mVolts)*	*Correction factor*	*Concentration*[#] *(mM)*	*Yield (%)*
40 kU L^{-1} D-acylase	11.2	0.92	146.67	48.9
40 kU L^{-1} D-acylase + **60** kU L^{-1} NAAAR G291D F323Y	15.8	1.08	243.20	81.1
40 kU L^{-1} D-acylase + **120** kU L^{-1} NAAAR G291D F323Y	20.7	1.01	297.12	99.0

[#] Concentration = (peak area * correction factor) / 0.0703

RP-HPLC analysis of *N*-dansyl-D-serine was performed with an isocratic elution (80:20 H_2O:ACN (0.01% TFA)). Un-reacted dansyl-Cl and hydrolysis product dansyl-OH, eluted at 1.5 mins. *N*-dansyl-D-serine eluted at 2.0 mins and *N*-dansyl-DL-phenylalanine eluted at 14.75 mins (**Figure 81**). A standard curve of *N*-dansyl-D-serine was generated by dansylation of D-serine standards (**Figure 82**).

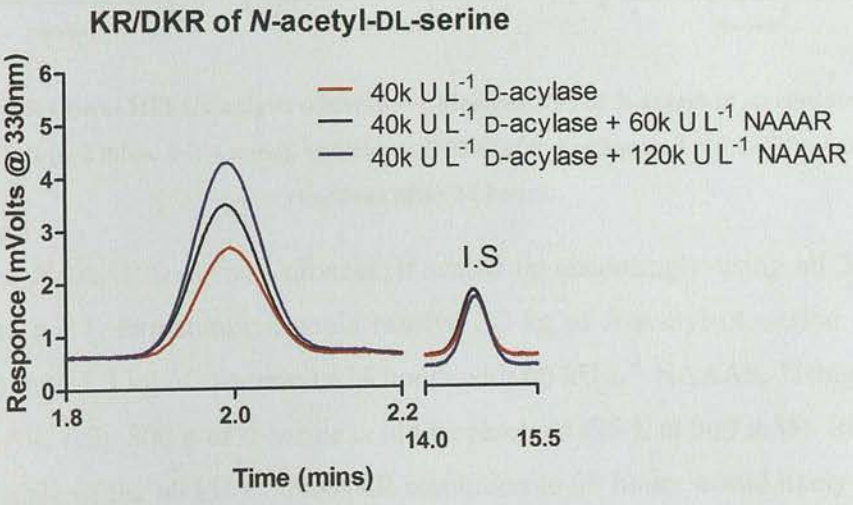


Figure 81. HPLC analysis (column = C8 Luna) of *N*-dansyl-D-serine product from both D-acylase KR and NAAAR/D-acylase DKR.

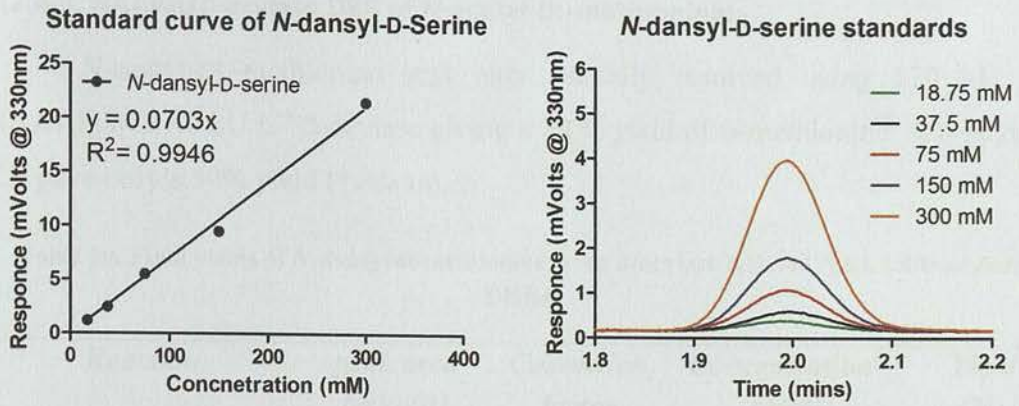


Figure 82. A Standard curve of *N*-dansyl-D-serine (after 1000 fold dilution) response at 330 nm, and HPLC traces of standards (column = C8 Luna).

Chiral HPLC was also carried out on the starting material to visualise the loss of the L-enantiomer upon addition of NAAAR. The hydrolysis of only *N*-acetyl-D-serine can be seen in the D-acylase KR, and hydrolysis of both enantiomers in the NAAAR/D-acylase DKR (**Figure 83**).

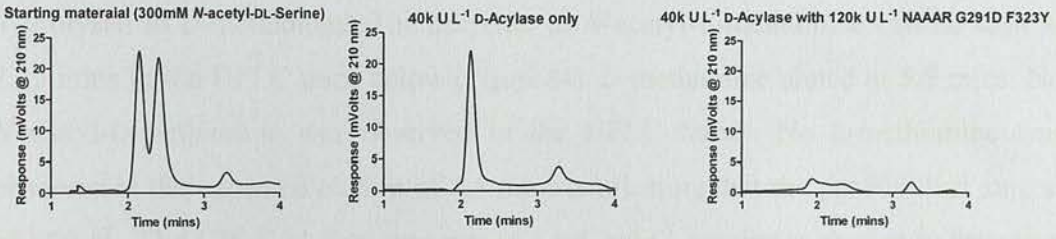


Figure 83. Chiral HPLC analysis (column = Chirobiotic T) of *N*-acetyl-DL-serine starting materials (L-2 mins, D-2.4 mins), and chiral HPLC of D-Acylase and NAAAR/D-Acylase reactions after 24 hours.

A coupled NAAAR/D-acylase process, if scaled up accordingly using all 3 million units from a 1 L fermentation, could resolve 2.2 kg of *N*-acetyl-DL-serine (50 L at 300 mM) in to 1.3 kg of D-serine in 24 hours with 60 kU L^{-1} NAAAR. Using 120 kU L^{-1} NAAAR, only 800 g of D-serine could be resolved (25 L at 300 mM). Increasing the time scale of the 60 kU L^{-1} NAAAR resolution to 30 hours would likely take the reaction to completion and allow 1.6 kg of D-serine to be resolved with 99% yield. Of this, 800 g would come from racemisation of *N*-acetyl-L-serine before hydrolysis.

5.2.4.3, NAAAR/D-acylase DKR of *N*-acetyl-DL-methionine:

N-acetyl-DL-methionine was only partially resolved using 120 kU L^{-1} NAAAR with 40 kU L^{-1} D-acylase giving a 79 % yield of D-methionine. A D-acylase KR gave only a 50% yield (**Table 18**).

Table 18. Final yields of *N*-dansyl-D-methionine from D-acylase KR and NAAAR/D-acylase DKRs.

Reaction	*peak area (mVolts)*	*Correction factor*	*Concentration*[#] *(mM)*	*Yield (%)*
40 kU L^{-1} D-acylase	358	1.00	149.2	49.7
40 kU L^{-1} D-acylase + 120 kU L^{-1} NAAAR G291D F323Y	585	0.97	236.4	78.8

[#] Concentration = (peak area * correction factor) / 0.0703

After 24 hours, only 60 % of *N*-acetyl-L-methionine had been racemized and hydrolysed to D-methionine. The decrease in *N*-acetyl-L-methionine can be seen at 2.25 mins in the HPLC trace below (**Figure 84**). D-methionine eluted at 5.5 mins. No *N*-acetyl-D-methionine was observed in the HPLC traces. No L-methionine was observed at the expected elution of 4.5 mins confirming that the product had an e.e. of >99 %. The HPLC elution was run at 1 mL min^{-1} causing a change in retention times compared to earlier HPLC work (See section 3.2.4.2).

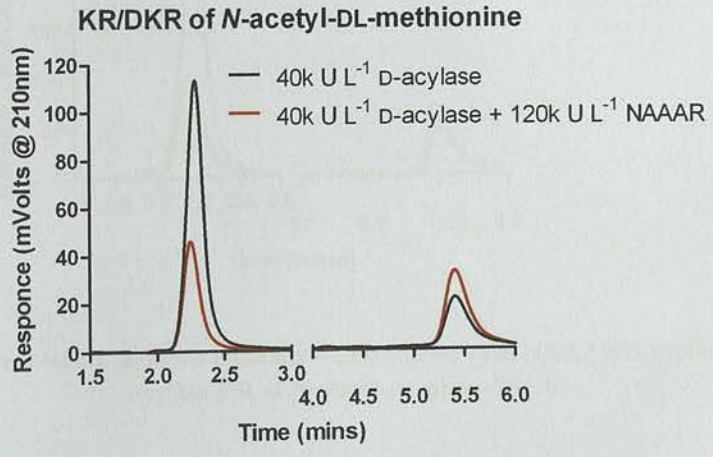


Figure 84. Chiral HPLC analysis (column = Chirobiotic T) of NAAAR/D-acylase DKR and D-acylase KR of *N*-acetyl-DL-methionine.

5.3.4.4, NAAAR/D-acylase DKR of *N*-acetyl-DL-phenylglycine:

NAAAR/D-acylase DKRs and D-acylase KRs of N-acetyl-DL-phenylglycine were attempted as this appeared to be a highly active substrate in earlier assays. This would also be an excellent industrial substrate as phenylglycines are useful synthetic building blocks and are also known to be potent ligands for Glutamate receptors in the central nervous system[271]. However, due to the highly insoluble nature of the D-phenylglycine product it proved difficult to effectively carry out these processes. The reactions would inevitably be almost entirely solid after 24 hours. Analysis of this solid showed the reaction was incomplete (~40% complete for D-acylase KRs, and ~70% for NAAAR/D-acylase DKRs). HPLC analysis of the NAAAR reaction indicated that although *N*-acetyl-L-phenylglycine concentration was dropping by 30%, D-phenylglycine was only increasing by 10%. This would suggest that some D-phenylglycine was being lost during the work up and skewing the final yield. Even using only the soluble *N*-acetyl-L-phenylglycine peak (elution time = 2.2 mins) to estimate final yield, the process would reach only ~80% yield, however, the reaction is possibly failing to reach 100% completion due to the formation of insoluble product. No L-phenylglycine was measured during analysis (elution time = 6.6 mins) giving the product an e.e. >99% (**Figure 85**).

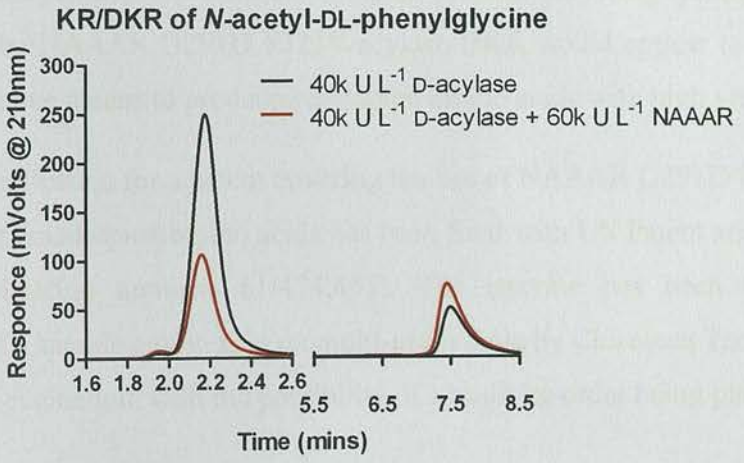


Figure 85. Chiral HPLC analysis (column = Chirobiotic T) of NAAAR/D-acylase DKR and D-acylase KR of *N*-acetyl-DL-phenylglycine.

5.3, Conclusions:

The industrial promise of NAAAR G291D F323Y has been successfully shown by small scale NAAAR/D-acylase DKRs of several *N*-acetylated amino acids. These reactions were conducted in the presence of 300 mM substrate over 24 hours –

typical industrial conditions. The final yield of D-amino acids in the NAAAR G291D F323Y/D-acylase DKRs was pushed towards 100%, a large improvement over the previous 50% maximum that was possible with a D-acylase KR. The e.e. of final product was measured by chiral HPLC to be >99%. The most active substrate, *N*-acetyl-DL-allylglycine, could be theoretically resolved on a multi kg scale using the total NAAAR biocatalyst from a 1 L fermentation and an appropriate acylase. This could possibly be a very effective method of producing enantiopure allylglycine. *N*-acetyl-DL-phenylglycine had shown promise in earlier assays but the resolution processes proved to be difficult to carry out and analyse due to the highly insoluble nature of the D-phenylglycine product.

NAAAR G291D F323Y has been shown to be a stable enzyme that can be produced by fermentation in high volume and recovered in reasonable yield (30% after DSP). Further optimisation of this and the DSP may be more beneficial than further mutagenesis for improving the cost efficiency of using NAAAR. Using a specialised fermentation host, such as *Pseudomonas fluorescens* would likely reduce the operating cost even further. Importantly for industrial processes, NAAAR has been shown to be stable at elevated temperatures for several hours and retains >90% activity over a range of 40 to 60°C. The lack of organic cofactor in both NAAAR and D/L-acylase removes the need for complex cofactor recycling systems which add extra cost. The NAAAR G291D F323Y/acylase DKR would appear to be a simple and cost effective means to produce enatiopure amino acids with high yields and e.e.

An application for a patent covering the use of NAAAR G291D F323Y in the production of enantiopure amino acids has been filed with US Patent and Trademark Office (Application number: 61/474,455). The enzyme has been used in the production of a sample amino acid on multi-gram scale by Chirotech Technology Ltd for customer evaluation, with the possibility of a multi kg order being placed.

5.4, Future work:

Future work on the industrial use of NAAAR will involve scale up and testing the reactions displayed here. The DKR of *N*-acetyl-DL-allylglycine looks like it may be the most successful biotransformation to come out of this work. The activity towards this substrate is pleasing as allylglycine is a precursor for pipeloic acid which is a chiral building block for Palinavir, a HIV protease inhibitor[272]. *N*-

acetyl-phenylglycines also look like they would be good substrates but further work and expertise is needed to test these adequately.

The D-acylase coupled to NAAAR has its highest activity with *N*-acetylated amino acids which is ideal for NAAAR. However, even though the L-acylase will accept *N*-acetylated amino acids, it has higher activity with *N*-benzoyl derivatives[86]. Although NAAAR is reported to have low, but measurable activity with *N*-benzoyl-amino acids, no activity was measured with *N*-benzoyl-D-phenylalanine when tested. Directed evolution could be applied to NAAAR using *N*-benzoyl-D-methionine (assuming SET21 can hydrolyse *N*-benzoyl amino acids, otherwise it will need complemented with L-acylase) as a source of L-methionine. If successful, this novel *N*-benzoyl-amino acid racemase would be an ideal enzyme to improve the yield of L-acylase resolutions of *N*-benzoyl amino acids.

Chapter 6, Experimental methods and materials:

6.1, Materials and reagents:

All general materials and reagents were purchased from Sigma-Aldrich or Fisher unless otherwise stated. Amino acid substrates were purchased from Sigma-Aldrich, Alfa Aesar, or provided by Chirotech Technology Ltd. All competent cells were purchased from Invitrogen or Promega. All oligos were purchased from Sigma-Genosys.

The SET21 cell line and pTTQ18 NAAAR WT plasmid were provided by Dr Ian Fotheringham (Ingenza Ltd).

6.2, Cell lines:

Cell line (supplier)	*Genotype*	*Application*
DH5α (Invitrogen)	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ^-	Plasmid propagation/plasmid cloning
Top10 (Invitrogen)	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ^-	Plasmid cloning
JM109 (Promega)	endA1 glnV44 thi-1 relA1 gyrA96 recA1 $mcrB^+$ Δ(lac-proAB) e14-[F' traD36 $proAB^+$ $lacI^q$ lacZΔM15] hsdR17($r_K^-m_K^+$)	pGEM cloning
BL21(DE3) (Novagen)	F^- ompT gal dcm lon $hsdS_B(r_B^-$ $m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Protein expression
SET21 (Ingenza)	Δ(araD-araB)567, ΔlacZ4787(::rrnB-4), $lacI^+$, λ^-, rpoS396(Am), rph-1, Δ (rhaD-RhaB)568, rrnB-4, hsdR514, ΔmetB, ΔdadA..dadX	NAAAR selection host
SET21(DE3) (this work)	As SET21 plus λ(DE3)	T7 promoter compatible NAAAR selection host

6.3, Microbiology reagents:

6.3.1, Agar:

All agars were dissolved in ddH_2O before being autoclaved (15 mins at 121°C).

LB agar: 35 g L^{-1} LB agar (Sigma).

S-gal agar: 38 g L^{-1} S-gal/LB agar mix (Invitrogen).

6.3.2, Broth:

All broths were dissolved in ddH_2O before being autoclaved (15 mins at 121°C).

LB broth: 10 g L^{-1} peptone, 5 g L^{-1} yeast extract, 10 g L^{-1} NaCl.
SOC broth: 10 g L^{-1} peptone, 5 g L^{-1} yeast extract, 0.88 g L^{-1} NaCl, 0.185 g L^{-1} KCl, 2.3 g L^{-1} $MgCl_2$, 1.36 g L^{-1} $MgSO_4$, 3.6 g L^{-1} glucose.

6.4, General techniques:

6.4.1, Heat shock transformation:

Chemical competent cells were mixed with 2 µL of chilled DNA and incubated for 20 minutes on ice. After a 45 second heat shock at 42°C, cells were mixed with 80 µL SOC and incubated at 37°C for 1 hour with agitation. 50 µL of this was spread on an LB agar plate supplemented with appropriate antibiotics. For cloning steps, *e.g.* ligations or mutagenesis, the entire transformation mixture was spread.

6.4.2, Electroporation:

Electrocompetent cells were mixed with 2 µL of DNA in a 0.2 cm gap electrocuvette (BioRad) and immediately electroporated with a single pulse of 2.5 kV, set to a capacitance of 25 µF and a resistance of 200 Ω. Cells were immediately resuspended in 1 mL of SOC preheated to 37°C. This was incubated at 37°C with agitation before 50 µL was spread on an LB agar plate supplemented with appropriate antibiotics. For SET21 selection assays, the 1 mL was washed repeatedly before spreading to remove exogenous L-methionine (See section 6.5.5).

6.4.3, Plasmid DNA isolation:

Plasmid DNA was isolated from a 5 mL LB broth overnight culture (supplemented with either 30 µg mL^{-1} kanamycin or 100 µg mL^{-1} ampicillin) using a QIAprep Spin Miniprep Kit (Qiagen) kit following the manufacturers protocol. DNA was eluted in 100 µL ddH_2O and stored at -20°C.

6.4.4, Big Dye sequencing method:

Plasmid DNA was sequenced using the Sanger dideoxy chain-termination method. Two sequencing PCRs were performed, one with the forward sequencing primer, and one with the reverse as follows:

5 μL plasmid

2 μL Terminator 3.1 buffer

2 μL Big Dye

1 μL primer

The following thermal cycle was used: 94°C for 30s, (94°C for 30s, 50°C for 20s, 60°C for 20s) x24. The DNA was sequenced using an ABI prisim 377 DNA sequencer using the Sanger dideoxy chain termination method by the Genepool at the Ashworth Laboratories. Sequencing primers used were standard pET forward/pET reverse, pGEM forward/pGEM reverse, M13for and pTTQ18rev (5'-GCC AGT GCC AAG CTT CTA-3')

6.5, SET21 Microbiology:

6.5.1, SET21 host growth conditions:

Non-transformed SET21 cells (from a 20% glycerol stock) were washed three times with sterile ddH_2O and streaked on replica DMA plates (with no antibiotic selection) containing either; 2 mM L-methionine, 2 mM D-methionine, 2 mM *N*-acetyl-L-methionine, or 2 mM *N*-acetyl-D-methionine. SET21(DE3) transformed with pET20b NAAAR WT cells (from overnight LB broth (100 μg mL^{-1} ampicillin) cultures) were washed three times with sterile ddH_2O and streaked on replica DMA plates (with no antibiotic selection) containing either; 2 mM L-methionine, 2 mM D-methionine, 2 mM *N*-acetyl-L-methionine, or 2 mM *N*-acetyl-D-methionine Plates contained 2 mg mL^{-1} 2,3,5-triphenyltetrazolium chloride (TTC) as a growth contrast agent. Plates were incubated at 37°C for 24 hours to allow for cell growth.

6.5.2, SET21(DE3) lysogenization:

SET21 were made λ(DE3) lysogens using a commercial kit and the manufacturer's instructions (Merck). Lysogenation was confirmed by over expression of NAAAR from pET28a NAAAR WT.

6.5.3, Minimal media:

6.5.3.1, M9 minimal agar:

M9 salts (10x):

$Na_2HPO_4.7H_2O$	128 g L^{-1}
KH_2PO_4	30 g L^{-1}

NaCl	5 g L^{-1}
NH_4Cl	10 g L^{-1}

Autoclaved for 15 minutes at 121°C.

1 M $MgSO_4$ – filter sterilised.

1 M $CaCl_2$ – filter sterilised.

20% glucose – filter sterilised.

1 M IPTG – filter sterilised.

60 mM $CoCl_2$ – filter sterilised.

100 mM methionine derivative – filter sterilised.

No trace metal or vitamin solution was added to minimal media.

To 270 mL of autoclaved agar number 1, the following were added:

720 µL $MgSO_4$

36 µL $CaCl_2$

7.2 mL 20% Glucose

36 mL M9 Salts (10x)

300 - 600 µL appropriate antibiotics

300 µL IPTG

1.5 mL $CoCl_2$

Methionine derivative added as required.

6.5.3.2, Davis minimal agar:

26.6 g L^{-1} Davis Minimal agar (autoclaved for 15 minutes at 121°C).

100 mM methionine derivative – filter sterilised.

To 100 mL of autoclaved Davis minimal agar, the following were added:

100 - 200 µL appropriate antibiotics.

Methionine derivative added as required.

This was then aliquoted as 25 mL into sterile Petri dishes.

6.5.3.3, Davis minimal broth:

11.6 g L^{-1} Davis Minimal broth (autoclaved for 15 minutes at 121°C).

100 - 200 µL appropriate antibiotics.

100 mM methionine derivative – filter sterilised.

To 100 mL of autoclaved Davis minimal broth, the following were added:

100 - 200 µL appropriate antibiotics.
Methionine derivative added as required.

6.5.4, SET21(DE3) cell competency:

6.5.4.1, $CaCl_2$ chemical competent cells:

SET21 and SET21(DE3) were made chemically competent by first streaking a strain glycerol stock on LB agar (30 µg mL^{-1} chloramphenicol) and picking a single colony into 5 mL LB broth (30 µg mL^{-1} chloramphenicol). This was incubated overnight at 37°C before back dilution into 100 mL LB broth (30 µg mL^{-1} chloramphenicol) to an OD_{600} of 0.1. This was allowed to grow at 37°C with agitation until the OD_{600} reached 0.5, at which point the culture was chilled on ice for 60 minutes. All following steps were performed in a cold room at 4°C with a chilled centrifuge and ice cold buffers. Cells were pelleted (*2,000 g*, 15 mins, 4°C) and resuspended with 100 mL of autoclaved 100 mM $CaCl_2$. These were incubated on ice for 30 minutes followed by another centrifugation (*2,000 g*, 15 mins, 4°C) and resuspension with 10 mL of 100 mM $CaCl_2$. This step was repeated before a final centrifugation (*2,000 g*, 15 mins, 4°C) and resuspension in 2 mL 100 mM $CaCl_2$ 15% glycerol. Cells were stored as 200 µL aliquots at -80°C.

6.5.4.2, Electrocompetent cells:

SET21 and SET21(DE3) were made electrocompetent by first streaking a glycerol on LB agar (30 µg mL^{-1} chloramphenicol) and picking a single colony into 5 mL LB broth (minus NaCl) (30 µg mL^{-1} chloramphenicol). This was incubated overnight at 37°C before back dilution into 100 mL LB broth (minus NaCl) (30 µg mL^{-1} chloramphenicol) to an OD_{600} of 0.1. This was allowed to grow at 37°C until the OD_{600} reached 0.5-0.6, at which point the culture was chilled on ice for 60 minutes. All following steps were performed in a cold room at 4°C with a chilled centrifuge and ice cold buffers. Cells were pelleted (*2,000 g*, 15 mins, 4°C) and resuspended with 25 mL autoclaved ddH_2O. These were incubated on ice for 30 minutes followed by another centrifugation (*2,000 g*, 15 mins, 4°C) and resuspension with 25 mL ddH_2O. Cells were pelleted again (*2,000 g*, 15 mins, 4°C) and resuspension in 5 mL 20% glycerol, before a final centrifugation (*2,000 g*, 15 mins, 4°C) and resuspension in 500 µL 20% glycerol, 150 mM trehalose. Cells were stored as 50 µL aliquots at -80°C.

6.5.4.3, Calculating transformation efficiency:

Both electro and chemically competent cells were measured for transformation efficiency by transforming with 0.01 ng of pUC18 (Invitrogen) using the standard transformation protocols. After 1 hour of phenolic outgrowth, four LB agar plates (100 μg mL^{-1} ampicillin) were spread with 10% of cells with no dilution, 1/10 dilution, 1/100 dilution, or 1/1000 dilution. The dilutions were performed as the neat and 1/10 dilution typically gave $>10^{2-3}$ colonies making counting impossible. Transformation efficiency, in c.f.u. per μg pUC18, was calculated using *Eqn (5)*.

Eqn (5), $$Tranformation\ efficiency = \frac{Number\ of\ colonies}{\mu g\ of\ pUC18\ *\ \%\ dilution}$$

6.5.5, SET21 genetic selection:

6.5.5.1, SET21 pTTQ18 G293x saturation mutagenesis selection:

SET21 cells electroporated with a mutagenic the pTTQ18 NAAAR G293x library were incubated at 37°C with agitation for 1 hour and then washed three times with 0.2 μm filtered ddH_2O to remove exogenous media containing L-methionine. The final resuspension was in 100 μL ddH_2O. 95 μL of this was spread on Davis minimal agar (100 μg mL^{-1} ampicillin, 30 μg mL^{-1} chloramphenicol) supplemented with 2.5 mM *N*-acetyl-D-methionine for selection, and the remaining 5 μL was spread on LB agar (100 μg mL^{-1} ampicillin, 30 μg mL^{-1} chloramphenicol) to check cloning efficiency. Colony size was judged by eye and those thought to be the largest on a selection plate were re-streaked onto a replica Davis minimal agar plate (100 μg mL^{-1} ampicillin, 30 μg mL^{-1} chloramphenicol, 2.5 mM *N*-acetyl-D-methionine). These were incubated at 37°C overnight and growth compared to a SET21 pTTQ18 NAAAR G293E standard. Those colonies that appeared to outgrow the parental standard were picked into 5 mL LB broth overnight cultures (100 μg mL^{-1} ampicillin, 30 μg mL^{-1} chloramphenicol). The plasmid DNA from these was isolated and sequenced using the Big dye method (See section 6.4.4).

6.5.5.2, SET21(DE3) pET20b G291D epPCR selection:

SET21(DE3) cells electroporated with a pET20b NAAAR G291D epPCR mutagenic library were incubated at 37°C with agitation for 1 hour and then washed three times with 0.2 μm filtered ddH_2O to remove exogenous media containing L-methionine. The final resuspension was in 100 μL. 95 μL of this was spread on Davis

minimal agar (100 μg mL $^{-1}$ ampicillin, 30 μg mL^{-1} chloramphenicol, 500 μM *N*-acetyl-D-methionine) and the remaining 5 μL was spread on LB agar (100 μg mL^{-1} ampicillin) to check cloning efficiency. Colony size was judged by eye and those thought to be the largest on a selection plate were picked into LB broth (100 μg mL^{-1} ampicillin, 30 μg mL^{-1} chloramphenicol). These cultures were incubated at 37°C overnight with agitation before being sub-cultured to an OD_{600} ~0.1 in Davis minimal broth (100 μg mL^{-1} ampicillin, 30 μg mL^{-1} chloramphenicol, 500 μM *N*-acetyl-D-methionine). The OD_{600} of these culture was monitored over the next 8 hours at 37°C and compared to growth of SET21(DE3) pET20b G291D NAAAR grown in the same media.

6.6, Cloning of pET28a and pET20b NAAAR:

pET28a and pET20b NAAAR WT was cloned using the WT NAAAR gene from pTTQ18 NAAAR WT with primers *NAAAR NdeI F* that encoded an 5'- NdeI site and *NAAAR HindIII R* encoding a 3'- HindIII site. The restriction sites are underlined and tha start codon mutation shown below in bold. The reaction contained:

125ng *NAAAR NdeI F* 5'-CAT **A**TG AAA CTC AGC GGT GTG GAA CTG-3'
125 ng *NAAAR HindIII* R 5'-AAG CTT CTA CGA ACC GAT CCA-3'
50 ng pTTQ18 NAAAR WT
5 μL dNTP (50x, 200 μM final concentration of each)
2 μL $MgSO_4$ (25x, 1 mM final concentration)
5 μL *KOD* polymerase buffer (10x)
1 μL *KOD* polymerase (1 U μL^{-1})
Up to 50 μL with ddH_2O

The following thermal cycle was used: 95°C for 5 mins, (95°C for 1 min, 50°C for 1 min, 72°C for 2 mins) x30, 72°C for 2 mins. A' overhangs were added by adding 1 Pure Taq Ready to Go bead (GE healthcare) to the finished PCR reaction and using the following thermal cycle: 95°C for 5 minutes, 72°C for 10 minutes. The PCR product (1116 bp) was gel purified using the standard protocol found in the Qiagen QAIquick® Gel extraction kit. This was then ligated into pGEM-Teasy using the following protocol:

1 μL pGEM-Teasy plasmid

3 μL NAAAR WT PCR product

1 μL T4 DNA ligase

5 μL Rapid ligation buffer (2x)

The reaction was left at 4°C overnight before being used to transform JM109 using the standard heat shock transformation protocol. Transformants were spread on S-gal LB agar plates (100 μg mL^{-1} ampicillin) and incubated at 37°C overnight. White colonies were selected against black/blue colonies and used to culture 5 mL LB broth (100 μg mL^{-1} ampicillin). Cultures were incubated at 37°C overnight and the plasmid DNA isolated. The presence of an insert was checked by a restriction digest with *EcoRI* and *HindIII*. Digestion products were run on 1% agarose gel. Plasmids which gave the correct bands (double cut – 1120bp & 3000bp; single cut – 4120bp) were checked by sequencing with the Big Dye method (See section 6.4.4).

Sequence confirmed pGEM T-easy NAAAR WT plasmids were digested with *NdeI* and *HindIII* to give a sticky ended fragment which was ligated as above, into pET28a or pET20b cut with the same restriction enzymes. The transformation was performed in DH5α and spread on LB agar (30 μg mL^{-1} kanamycin for pET28a or 100 μg mL^{-1} ampicillin for pET20b). Transformants were picked into 5 mL LB (30 μg mL^{-1} kanamycin for pET28a or 100 μg mL^{-1} ampicillin for pET20b) and grown overnight at 37°C. The plasmid DNA was isolated from these cultures and checked for the presence of an insert with an *NdeI/HindIII* restriction digest. Plasmids with confirmed inserts were stored at -20°C. This generated the plasmids pET28a NAAAR WT and pET20b NAAAR WT. To generate pET28a NAAAR G291E and pET20b NAAAR G291E, this procedure was repeated with pTTQ NAAAR G293E as the initial PCR template (See section 6.7.1).

6.7, Mutagenesis:

6.7.1, Site directed mutagenesis of pTTQ18 NAAAR G293E:

pTTQ18 NAAAR WT was used as a template for SDM at position G293 using primers *G293E for* and *G293E rev* which encode a G293E mutation. The introduced DNA mutation is shown below in bold. The reaction contained:

125 ng *G291E for* – 5'-GG TGC GGC G**A**G ATG ATC GAG ACC G-3'

125 ng *G291E rev* – 5'-CC GGT CTC GAT CAT C**T**C GCC GCA C-3'

50 ng pTTQ18 NAAAR WT

1 µL dNTPS (50x, 200 µM final concentration of each)

5 µL *Pfu* reaction buffer (10x)

1 µL *Pfu* turbo (2.5 U μL^{-1})

Up to 50 µL with ddH_2O

The following thermal cycle was used: 94°C for 30s, (94°C for 30s, 50°C for 20s, 72°C for 2 mins) x24. 10 µL of PCR product was digested for 5 hour at 37°C with 1 µL *DpnI* (1 U μL^{-1}) to remove the methylated parental plasmid. 2 µl of the digested PCR product was then used to transform 50 µl of Top10 cells by heat shock and spread on LB agar (100 µg mL^{-1} ampicillin). Several transformants were used to culture separate 5 mL LB (100 µg mL^{-1} ampicillin) overnights at 37°C. The plasmid DNA from these cultures was isolated and sequenced using the Big Dye method (See section 6.4.4).

6.7.2, Saturation mutagenesis of pTTQ18 NAAAR G293:

Site directed saturation mutagenesis of G293 was performed on pTTQ18 NAAAR WT using primers *G293x for* and *G293x rev*. The degenerate NNK codon is shown below in bold. The reaction contained:

125 ng *G293x for*: 5'-C CCG GTG TGG TGG TGC GGC **NNK** ATG-3'

125 ng *G293x rev*: 5'-CC GGT CTC GAT CAT **MNN** GCC G-3'

50 ng pTTQ18 NAAAR WT

1 µL dNTPS (50x, 200 µM final concentration of each)

5 µL *Pfu* reaction buffer (10x)

1 µL *Pfu* turbo (2.5 U μL^{-1})

Up to 50 µL with ddH_2O

The following thermal cycle was used: 94°C for 30s, (94°C for 30s, 50°C for 20s, 72°C for 2 mins) x24. 10 µL of PCR product was digested for 5 hour at 37°C with 1 µL *DpnI* (1 U μL^{-1}) to remove the methylated parental plasmid. This product was used to transform SET21 for selection of improved variant NAAARs (See section 6.5.5.1).

6.7.3, Error prone PCR of pET20b NAAAR G291D:

A mutagenic pET20b G291D NAAAR library was constructed using *pET20b Fm* and *pET20b Rm* which terminated on the 3' end with either the ATG start codon

or CTA stop codon reverse compliment. The start and stop codons are shown below in bold and the restriction sites underlined. The reaction contained:

250 ng *pET20b Fm* 5'-GTT TAA CTT TAA GAA GGA GAT ATA CAT **ATG**-3'
250 ng *pET20b Rm* 5'-GTG CTC GAG TGC GGC CGC AAG CTT **CTA**-3'
500 ng pET20b NAAAR G291D
1 µL dNTPS (50x, 200 µM final concentration of each)
5 µL *mutazyme* reaction buffer (10x)
1 µL *mutazyme* (2.5 U μL^{-1})
Up to 50 µL with ddH_2O

The following thermal cycle was used: 98°C for 2 mins, (98°C for 30s, 55°C for 30s, 72°C for 2 mins) x24, 72°C for 10 mins. 2 µL of this PCR product was used in a second, high fidelity PCR to increase the amount of mutagenic PCR product. The reaction contained:

250 ng *pET20b Fm* 5'-GTT TAA CTT TAA GAA GGA GAT ATA CAT **ATG**-3'
250 ng *pET20b Rm* 5'-GTG CTC GAG TGC GGC CGC AAG CTT **CTA**-3'
1 µL pET20b NAAAR G291D epPCR product
1 µL dNTPS (50x, 200 µM final concentration of each)
5 µL *Pfu turbo* reaction buffer (10x)
1 µL *Pfu turbo* (2.5 U μL^{-1})
Up to 50 µL with ddH_2O

The following thermal cycle was used: 98°C for 2 mins, (98°C for 30s, 55°C for 30s, 72°C for 2 mins) x24, 72°C for 10 mins. This PCR product (1152 bp) was gel purified before cloning into a pET20b plasmid using a MEGAWHOP PCR containing:

>300 ng gel purified high fidelity epPCR product (megaprimer)
50 ng pET20b NAAAR WT
1 µL dNTPS (50x, 200 µM final concentration of each)
5 µL *Pfu turbo* reaction buffer (10x)
1 µL *Pfu turbo* (2.5 U μL^{-1})
Up to 50 µL with ddH_2O

The following thermal cycle was used: 98°C for 2 mins, (98°C for 20s, 55°C for 20s, 68°C for 2 mins) x25. 10 µL of PCR product was digested for 5 hour at 37°C with 1

μL *DpnI* (1 U μL^{-1}) to remove the methylated parental plasmid. This product was used to transform SET21(DE3) for selection of improved variant NAAARs (See section 6.5.5.2).

6.7.4, Saturation mutagenesis of pET20b NAAAR G291D F323:

Saturation mutagenesis of F323 was performed by creating a megaprimer from pET20b NAAAR G291D using primers *pET20b Fm* and *F323x rev*. This megaprimer was incorporated into pET20b NAAAR G291D using a MEGAWHOP PCR. The degenerate NNK codon is shown below in bold, with the NdeI restriction site underlined.

1 μg *pET20b Fm*: 5'- GTT TAA CTT TAA GAA GGA GAT ATA CAT ATG -3'
1 μg *F323x rev*: 5'- GGT TTT GTG **MNN** CCG GTC C -3'
50 ng pET20b NAAAR G291D
1 μL dNTPS (50x, 200 μM final concentration of each)
5 μL *Pfu turbo* reaction buffer (10x)
1 μL *Pfu turbo* (2.5 U μL^{-1})
Up to 50 μL with ddH_2O

The following thermal cycle was used: 98°C for 30s, (98°C for 30s, 55°C for 30s, 72°C for 2 mins) x40) 72°C for 10 mins. This PCR product was gel purified before cloning into a pET20b plasmid using a MEGAWHOP PCR containing:

>300 ng gel purified high fidelity G291x PCR product (megaprimer)
50 ng pET20b NAAAR G291D
1 μL dNTPS (50x, 200 μM final concentration of each)
5 μL *Pfu turbo* reaction buffer (10x)
1 μL *Pfu turbo* (2.5 U μL^{-1})
Up to 50 μL with ddH_2O

The following thermal cycle was used: 98°C for 2 mins, (98°C for 20s, 55°C for 20s, 68°C for 2 mins) x25. 10 μL of PCR product was digested for 5 hour at 37°C with 1 μL *DpnI* (1 U μL^{-1}) to remove the methylated parental plasmid. This product was used to transform BL21(DE3) for high-throughput, chiral HPLC screening of improved variant F323x NAAARs (See section 6.8).

6.8, High-throughput HPLC screening of pET20b NAAAR G291D F323x libraries:

2 μL of digested pET20b G291D F323x saturation mutagenesis product was used to transform BL21(DE3) and this spread on LB agar (100 μg mL^{-1} ampicillin). After an overnight incubation at 37°C, 96 random colonies were picked into 1 ml of LB broth (100 μg mL^{-1} ampicillin) in two 2.2 mL 48 well plates. These were incubated at 37°C overnight with agitation and a gas permeable sealed top. These cells were then used in a small scale biotransformation of *N*-acetyl-L-methionine to look for variants with improved racemase activity. The reaction was carried out by adding 10 μL cells to 90 μL reaction buffer (final concentrations upon addition of cells: 100 mM Tris-HCl (pH 8.0), 5 mM $CoCl_2$, 150 mM *N*-acetyl-L-methionine). These reactions, incubated at 60°C for 60 minutes, were carried out in 2.2 mL 96 well plates. The reactions were terminated with addition of 900 μL of 0.01% TEAA:MeOH (75:25), and diluted a further 100 fold with the same buffer. Reactions were then analysed by chiral HPLC[256] (See section 6.12). Conversion of *N*-acetyl-L-methionine to *N*-acetyl-D-methionine was used to assess enzyme activity. The plasmid DNA from the most active wells was isolated from the master growth plates and sequenced.

6.9, his-NAAAR and native NAAAR expression:

6.9.1, Expression of his-NAAAR:

WT and G291E hexa-histidine tagged NAAAR (his-NAAAR) was expressed in BL21(DE3) transformed with pET28a NAAAR WT/G291E. A single transformant from an LB agar plate (30 μg mL^{-1} kanamycin) was used to inoculate two 250 mL seed cultures of LB broth (30 μg mL^{-1} kanamycin). One of these of these was used to sub-culture 4.5 L of sterile LB broth (30 μg mL^{-1} kanamycin) to an OD_{600} ~0.1 and aliquoted into ten sterile 500 mL baffle flasks. These were incubated with agitation at 37°C until the OD_{600} reached 0.6, at which point expression was induced with 1 mM IPTG. Cells were harvested after 5 hours at 37°C (*5,000 g*, 20 mins, 4°C) and stored at -20°C.

6.9.2, Expression of native NAAAR:

Native NAAAR (WT and all variants) was expressed in BL21(DE3) transformed with pET20b expressing the NAAAR of interest. A single transformant

from an LB agar plate (100 μg mL^{-1} ampicillin) was used to inoculate 500 mL of sterile LB broth (100 μg mL^{-1} ampicillin). This was incubated at 37°C for 24 hours with no IPTG induction. Cells were then harvested (*5,000 g*, 20 mins, 4°C) and often used immediate for NAAAR purification (See section 6.10.2). If not used immediately, pellets were stored at -20°C.

6.10, his-NAAAR and native NAAAR purification:

6.10.1, Purification of his-NAAAR:

His-tagged NAAAR was purified using Ni-affinity chromatography, with all steps performed at 4°C or on ice. A BL21(DE3) cell pellet expressing his-NAAAR was re-suspended in 10 mL lysis buffer (50 mM Tris-HCl (pH 7.5), 20 mM imidazole, 1 complete mini EDTA free protease inhibitor tablet(Roche)) and lysed using 30 s on/30 s off sonication (Soniprep150). The resulting cell free extract was then clarified by centrifugation (*18,000 g*, 30 mins, 4°C), after which the supernatant was removed, 0.45 μm filtered, and gently mixed with 1 mL of Ni:NTA beads (Invitrogen) for 60 mins at 4°C. Ni:NTA beads were previously washed and equilibrated with lysis buffer.

The protein bound Ni:NTA beads were loaded into a 1 mL polypropylene column (Qiagen) using a peristaltic pump and eluted using a step wise gradient of imidazole (20 cv of 50 mM, 10 cv of 100 mM, and 10 cv of 250 mM). Fractions were tested for the presence of his-NAAAR by SDS-PAGE analysis. The 100 mM and 250 mM imidazole fractions, both of which usually contained his-NAAAR were dialysed three times against 1 L of 50 mM Tris-HCl (pH 7.5) with an 8 kDa MWCO membrane. Protein concentration was estimated from Abs_{280} measured on a Varian Cary 50 ($M_{r\ WT}$ = 41,678 g, $M_{r\ G291E}$ = 41,750 g, $\varepsilon_{WT,\ G291E}$ = 33,600 M^{-1} cm^{-1}). Protein yield was roughly 10 mg L^{-1} of growth media and judged to be at least 95% pure. Protein concentration was adjusted to 1 mg mL^{-1} and frozen as 20% glycerol at -20°C.

6.10.2, Purification of native NAAAR:

Native NAAAR was purified using Anion Exchange and Size Exclusion chromatography, all steps performed at 4°C or on ice. After 24 hours of expression, 500 mL of LB broth (100 μg mL^{-1} ampicillin) containing BL21(DE3) pET20b NAAAR was harvested (*6,000 g*, 20 mins, 4°C) and then resuspended in 10 mL lysis

buffer (50 mM Tris-HCl (pH 8.0) 100 mM NaCl, 2 mg mL^{-1} Lysozyme, 1 mini complete EDTA free protease inhibitor tablet (Roche) and left for 30 minutes on ice. Cells were lysed by sonication (30s on/ 30s off) and the cell free extract clarified by centrifugation (*18,000 g*, 60 mins, 4°C). The supernatant was removed and 0.45 µm filtered before loading onto a HiPrep 16/10 FF Q column (GE healthcare) equilibrated with 50 mM Tris-HCl (pH 8.0), 100 mM NaCl. NAAAR was then eluted using a 2 stage gradient (0 to 45% over 9 column volumes, 45 to 100% over 3 column volume) of 50 mM Tris-HCl (pH 8.0), 1 M NaCl. Fractions judged to contain the highest purity and concentration of NAAAR were pooled and concentrated to less than to 1 mL using a 30 kDa cut off spin filter (Vivaspin). The concentrated sample was then loaded on to an HiPrep 26/60 Sephacryl S-300 HR (GE healthcare) equilibrated with 50 mM Tris-HCl (pH 8.0), 100 mM NaCl. NAAAR was eluted over an isocratic elution at 1 mL min^{-1}. WT NAAAR, and all variants, eluted as octamers at roughly 135 mL. Protein concentration was calculated using Abs_{280} (Mr_{WT} = 39,407 g, Mr_{G291E} = 39,479 g, Mr_{G291D} = 39,463 g, $Mr_{G291D\ F323Y}$ = 39,481 g, $\varepsilon_{WT,\ G291E,\ G291D,\ G291D\ F323Y}$ = 33,600 M^{-1} cm^{-1}). Protein samples were submitted to SIRCAMS (The University of Edinburgh) at a concentration of 1 mg mL^{-1} in size exclusion elution buffer for ESI-MS analysis. The Sephacryl S-300 column was calibrated with Carbonic Anhydrase (M_r = 29 kDa), Conalbumin (75 kDa), Aldolase (158 kDa), and Thyroglobulin (669 kDa), all eluted in 50 mM Tris-HCl (pH 8.0), 100 mM NaCl. The void volume was found to be 92.101 mL with Blue Dextran (M_r >2,000 kDa) and the column volume was 320 mL. This gave *Eqn (6)* and *Eqn (7)* which were used to calculate molecular weight:

Eqn (6),
$$K_{av} = \frac{Elution\ volume - 92.101}{227.899}$$

Eqn (7),
$$Molecular\ weight = e\left(\frac{K_{av} - 1.8272}{-0.129}\right)$$

6.11, In vitro NAAAR assays:

Purified his-NAAAR and native NAAAR were both assayed using chiral HPLC. Purified enzyme was incubated with a single *N*-acetyl-amino acid enantiomer in the following reaction.

100 µL purified NAAAR (0.1 mg mL^{-1} final concentration)

100 µL $CoCl_2$ (10x, 5 mM final concentration)

100 µL Tris-HCl (10x, 50 mM final concentration)

500 µL *N*-acetyl amino acid in 100 mM Tris-HCl (pH 8.0) (2x)

200 µL ddH_2O

Substrates were made up in 100 mM Tris-HCl and the pH adjusted to 8.0 at 60°C. Substrate solutions were 2x stocks giving an overall final Tris-HCl concentration of 100 mM.

Reactions were terminated and diluted 20 fold with 50 mM HCl and boiled for 3 minutes to precipitate protein. These were clarified by centrifugation (*11,000 g*, 5 mins, rt) and the supernatant 0.45 µm filtered before analysis using an Agilent 1100 or 1200 HPLC system (Ingenza, Roslin Biocentre). HPLC conditions were as follows:

Column: Chirobiotic™ T (5 µm, 250 mm x 4.6 mm)
Flow rate: 0.5 mL min^{-1}
Mobile phase: 15 minute isocratic elution of 75:25 0.01% TEAA: MeOH. TEAA was a 1:1 (v/v) of glacial acetic acid and triethylamine.
Detection λ: 210 nm
Injection volume: 5 µL
Temperature: 40°C

Elution times: *N*-acetyl-L-methionine 4.5 mins, *N*-acetyl-D-methionine 6.1 mins, L-methionine 8.5 mins, and D-methionine 10.0 mins.

6.12, In vivo NAAAR assays:

Biotransformations were carried out by transforming BL21(DE3) cells with pET20b NAAAR WT and G291E, G291D, or G291D F323Y variants. These were used to inoculate 5 mL of LB (100 µg mL^{-1} ampicillin) and grown overnight at 37°C with agitation. No IPTG was added to these cultures. To generate a range of NAAAR concentrations these cells were diluted by factors of 1.0 (no dilution), 0.75, 0.5 and 0.25, and 10 µL of each diluted culture was incubated with 50 µL substrate (final concentration, 150 mM), 10 µL Tris-HCl (stock concentration, 500 mM, pH 8.0 at 60°C), 10 µL $CoCl_2$ (stock concentration, 50 mM) and 20 µL dH_2O (final volume 100 µL). The substrates were suspended in 100 mM Tris-HCl and the pH adjusted to 8.0 at 60°C. This gave a final concentration of 100 mM Tris-HCl in each reaction.

The reactions were incubated at 60°C for 1 hour with agitation in a 2.2 mL 96 deep well plate. The reactions were terminated and diluted with three 10 fold dilutions in 0.01% TEAA:MeOH (75:25). The enantiomeric excess was measured by chiral HPLC as follows:

Column: Chirobiotic™ T (5 μm, 250 mm x 4.6 mm)
Mobile phase: isocratic elution of 75:25 0.01% TEAA: MeOH. TEAA was a 1:1 (v/v) of glacial acetic acid and triethylamine.
Flow rate: 0.5 mL min^{-1}
Injection volume: 20 μL
Detection λ: 210 nm
Temperature: 40°C

Elution times: *N*-acetyl-L-methionine 4.4 mins, *N*-acetyl-D-methionine 6.9 mins, *N*-acetyl-L-phenylalanine 4.8 mins, *N*-acetyl-D-phenylalanine 18.9 mins, *N*-acetyl-(2-aminobutyric acid), *N*-acetyl-L-allylgylcine 4.2 mins, *N*-acetyl-D-allylgylcine 5.8 mins, *N*-acetyl-L-(4-fluorophenylglycine) 5.1 mins, *N*-acetyl-D-(4-fluorophenylglycine) 18.0 mins, and *N*-acetyl-L-phenylglycine 5.0 mins, *N*-acetyl-D-phenylglycine 18.5 mins. The column was calibrated with single enantiomers of *N*-acetyl derivatives.

Rate constants were calculated using *Eqn (1)*, *Eqn (2)*, and *Eqn (3)*, where D = proportion of D-enantiomer, and L = proportion of L-enantiomer:

Eqn (1),
$$2k_L = \ln \frac{\left(1+\frac{D}{L}\right)}{\left(1-\frac{D}{L}\right)}$$

Eqn (2),
$$2k_D = \ln \frac{\left(1+\frac{L}{D}\right)}{\left(1-\frac{L}{D}\right)}$$

Eqn (3),
$$k_{D/L} = \frac{slope}{2}$$

Specific activities were then calculated for 1 OD_{600} unit for each pET20b NAAAR BL21(DE3) strain using *Eqn (4)*:

Eqn (4), $$specific\ activity = \frac{\left((total\ \mu moles\ of\ substrate)k_{D/L}\right)}{OD}$$

6.13, Whole cell DKRs with NAAAR and an L-acylase:

6.13.1, Expression of NAAAR:

NAAAR WT, G291E, G291D, and G291D F323Y were expressed in BL21(DE3) from the appropriate pET20b plasmid in LB broth (100 μg mL^{-1} ampicillin) over 24 hours with no IPTG.

6.13.2, Expression of L-acylase:

L-acylase was expressed from a propriety plasmid (Chirotech Technology Ltd) under the control of a *lac* promoter. Expression was in auto-induction media (10 g L^{-1} peptone, 5 g L^{-1} yeast extract, 50 mM $(NH_4)_2SO_4$, 100 mM KH_2PO_4, 100 mM Na_2HPO_4, 0.5% glycerol, 0.05% glucose, 0.2% lactose, 1 mM $MgSO_4$, 100 μg mL^{-1} ampicillin, pH 7.5) for 24 hours at 37°C. No IPTG was added to this culture.

6.13.3, L-acylase kinetic resolution:

Kinetic resolution were started by adding 0.2 OD_{600} units of BL21(DE3) expressing L-acylase to a 5 mL reaction buffer containing 100 mM Tris-HCl (pH 8.0) 5 mM $CoCl_2$, 30 mM *N*-acetyl-DL-methionine. 50 μL samples were taken at regular time points to monitor reaction progress. These samples were terminated and diluted 20 fold in 50 mM HCl, before boiling for three minutes and being centrifuged (*11,000 g*, 5 mins, rt). The supernatant was then 0.45 μm filtered and analysed by chiral HPLC (See section 6.11).

6.13.4, NAAAR/L-acylase dynamic kinetic resolution:

A dynamic kinetic resolution was carried out exactly the same as the kinetic resolution above, but 0.2 OD_{600} units of BL21(DE3) expressing each NAAAR variant was also included. Samples were taken and analysed by chiral HPLC (See section 6.11).

6.14, Fermentation of BL21(DE3) pET20b NAAAR G291D F323Y:

The following fermentation was carried out at Chirotech Technology Ltd with the help of Dr Ian Taylor. BL21(DE3) were transformed with pET20b G291D F323Y NAAAR and spread on LB agar (100 μg mL^{-1} ampicillin). A single

transformant from this plate was used to inoculate a shake flask containing 50 mL of *E. coli* seed media (40 g L^{-1} Amisoy, 5 g L^{-1} HiYeast, 0.5 g L^{-1} $MgSO_4.7H_2O$, 0.5 g L^{-1} NaCl, 5 g L^{-1} KH_2PO_4, 100 μg mL^{-1} ampicillin, pH 7.0). This was cultured at 30°C for 8 hours before 2.5 mL was used to inoculate a 1 L fermentor containing 700 mL of *E. coli* fermentation media (15 g L^{-1} Amisoy, 15 g L^{-1} HiYeast, 4 g L^{-1} $(NH_4)_2SO_4$, 8 g L^{-1} KH_2PO_4, 7 g L^{-1} K_2HPO_4, 1 g L^{-1} $MgSO_4.7H_2O$, 1 mL L^{-1} trace elements solution (Dow), 2 mL L^{-1} antifoam, 50 mL L^{-1} glycerol, 100 μg mL^{-1} ampicillin, pH 7.0). After 19.5 hours, *E. coli* feed media (40 g L^{-1} Amisoy, 20 g L^{-1} HiYeast, 5 g L^{-1} $(NH_4)_2SO_4$, 1 g L^{-1} $MgSO_4.7H_2O$, 208 g L^{-1} glycerol, 1.25 mL L^{-1} trace elements solution, 100 μg mL^{-1} ampicillin, pH 7.0) was fed in at a rate of 14 mL $hour^{-1}$. After 40 hours at 30°C, the cells were harvested (*6,000 g*, 30 mins, 4°C) to give a wet weight of 90.8 g (16.2 g dry cells) from a total of 1.05 L media.

Samples were taken throughout the fermentation process to monitor NAAAR activity. Total NAAAR activity present in the fermentation broth and specific activities were calculated by testing the fermentation broth under standard NAAAR assay conditions (150 mM *N*-acetyl-L-methionine in 100 mM Tris-HCl (pH 8.0), 5 mM $CoCl_2$, for 1 hour at 60°C, in a reaction volume of 100 μL. The decrease in e.e. was measured as above using chiral HPLC (See section 6.12). The specific activity was determined by calculating the activity per OD_{600} of each fermentation sample. Total activity was calculated by measuring the total OD_{600} of the fermentation.

6.15, BL21(DE3) pET20b NAAAR G291D F323Y downstream processing (DSP):

10 g of fermented wet cells were resuspended in 50 mL of 50 mM Tris-HCl (pH 8.0), 100 mM NaCl and lysed with sonication (20 s on/ 20 s off, 15 mins, 2°C). This was clarified by centrifugation (*12,000 g*, 20 mins, 4°C) and the supernatant stored at -20°C. The cell free extract (50 mL) was assayed under standard NAAAR assay conditions (150 mM *N*-acetyl-L-methionine in 100 mM Tris-HCl (pH 8.0), 5 mM $CoCl_2$, for 1 hour at 60°C, in a reaction volume of 100 μL and found to contain 6615 (± 660) U mL^{-1}, where 1 U = the amount of enzyme required to racemise 1 μmole of *N*-acetyl-L-methionine in 60 minutes at 60°C. This reaction was analysed by chiral HPLC (See section 6.12).

6.16, Substrate profiling of NAAAR G291D F323Y cell free extract:

The CFE generated from the BL21(DE3) NAAAR G291D F323Y fermentation DSP was used to test substrate specificity with a variety of amino acids. The reactions were run for 1 hour at 60°C on a 100 µL scale. The reaction contained, 50 mM Tris-HCl (pH 8.0), 5 mM $CoCl_2$, 300 mM substrate in 100 mM Tris-HCl (pH 8.0), and an appropriate volume of NAAAR G291D F323Y CFE. The final concentration of Tris-HCl was 100 mM. The reaction was terminated with addition of 900 µL 0.01% TEAA:MeOH (75:25). This was then diluted a further 100 fold in the same buffer to give a final 1000 fold dilution of the reaction. The rate of racemisation was measured by chiral HPLC (See section 6.12).

The following amino acids were tested: *N*-acetyl-DL-methionine, *N*-acetyl-D-allylglycine, and *N*-acetyl-DL-phenylglycine, *N*-acetyl-D-(4-fluoro)phenylglycine, *N*-acetyl-DL-phenylalanine, *N*-acetyl-DL-(2-amino)butyric acid, *N*-acetyl-L-orthinine, *N*-acetyl-D-naphthylalanine, *N*-benzoyl-DL-phenylalanine.

6.17, Thermal studies of NAAAR G291D F323Y CFE:

6.17.1, Thermal stability of NAAAR G291D F323Y CFE:

5 µL NAAAR G291D F323Y CFE was incubated with 150 mM *N*-acetyl-L-methionine in reaction buffer (100 mM Tris-HCl, 5 mM $CoCl_2$, final reaction volume = 50 µL) in a 12 well PCR strip. This was then incubated in a PCR thermal cycler running a well temperature gradient increasing ~2.5°C per well (30 to 55°C) for 60 minutes before termination and 1000 fold dilution with HPLC running buffer 0.01% TEAA:MeOH (75:25). The samples were analysed by chiral HPLC (See section 6.12). This process was repeated with the thermal cycler running a well gradient between 55 and 80°C.

6.17.2, Thermal activity profiling of NAAAR G291D F323Y CFE:

NAAAR G291D F323Y CFE was incubated at 60°C for 5 hours, and tested for NAAAR activity under standard NAAAR assay conditions (150 mM *N*-acetyl-L-methionine in 100 mM Tris-HCl (pH 8.0), 5 mM $CoCl_2$, for 1 hour at 60°C, in a reaction volume of 100 µL. Samples were terminated and diluted 1000 fold in HPLC running buffer 0.01% TEAA:MeOH (75:25) and samples were analysed by chiral HPLC (See section 6.12).

6.18, NAAAR/D-acylase dynamic kinetic resolutions and D-acylase kinetic resolutions:

Reactions (1 mL) were prepared containing 100 mM Tris-HCl (pH 8.0), 5 mM $CoCl_2$, 300 mM *N*-acetyl-DL-amino acid (pH 8.0), 40 U D-acylase (Chirotech Technology Ltd), and 30-120 U NAAAR G291D F323Y (or buffer in place of this for kinetic resolutions). These reactions were incubated at 40°C for 24 hours with agitation. After this time, the *N*-acetyl-DL-serine and *N*-acetyl-DL-allylglycine reactions were terminated and diluted 10 fold in HPLC running buffer (75:25 H_2O:ACN) containing 10 mM DL-phenylalanine as an internal standard (I.S). This reaction was then diluted a further 10 fold in HPLC running buffer before a final 10 fold dilution in 50 mM sodium carbonate buffer (pH 9.8). To this 100 µL of 3 mg mL^{-1} 5-(dimethylamino)naphthalene-1-sulfonyl chloride (dansyl-Cl) in 100% acetone was added. This was incubated at 37°C for three hours before the reaction was terminated with 100 µL of glacial acetic acid. The conversion of *N*-acetyl-DL-allylglycine/*N*-acetyl-DL-serine to *N*-dansyl-D-allylglycine/*N*-dansyl-D-serine was measured by RP-HPLC using the following isocratic conditions:

Column: Luna (Phenomenex) C8(2) (5 µm, 150 mm x 4.6 mm)
Mobile phase (*N*-dansyl-serine): isocratic elution of 80:20 H_2O:MeOH (0.01% TFA)
Mobile phase (*N*-dansyl-allylglycine): isocratic elution of 75:25 H_2O:MeOH (0.01% TFA)
Flow rate: 2.0 mL min^{-1}
Detection λ: 330 nm
Injection volume: 50 µL
Temperature: rt

Elution times: *N*-dansyl-serine 2.0 mins, *N*-dansyl-allylglycine 4.15 mins, *N*-dansyl-phenylalanine 14.75 mins under *N*-dansyl-serine conditions and 6.7 mins under *N*-dansyl-allylglycine conditions.

N-acetyl-DL-serine, *N*-acetyl-DL-methionine, and *N*-acetyl-DL-phenylglycine reactions were terminated and diluted 1000 fold in HPLC running buffer. These were analysed by chiral HPLC using the following conditions.

Column: Chirobiotic™ T (5 µm, 250 mm x 4.6 mm)
Flow rate: 1 mL min^{-1}

Mobile phase: 15 minute isocratic elution of 75:25 0.01% TEAA: MeOH. TEAA was a 1:1 (v/v) of glacial acetic acid and triethylamine.

Detection λ: 210 nm

Injection volume: 50 µL

Temperature: rt

Elution times: *N*-acetyl-L-methionine 2.25 mins, *N*-acetyl-D-methionine 3.05 mins, *N*-acetyl-L-serine 2.0 mins, *N*-acetyl-D-serine 2.4 mins, *N*-acetyl-L-phenylglycine 2.2 mins, *N*-acetyl-D-phenylglycine 3.0 mins.

6.19, Melting temperature determination of NAAAR WT, G291E, G291D, and G291D F323Y:

The melting temperatures (T_m) of NAAAR WT, NAAAR G291E, NAAAR G291D and NAAAR G291D F323Y were determined using a BioRad iQ5 rtPCR Thermocycler at the Edinburgh Protein Production Facility. Each NAAAR variant was purified by Anion Exchange and Size Exclusion Chromatography (See section 6.10.1). Final NAAAR concentration was 0.5 mg mL^{-1}. T_m for apo- and holo-forms was determined without ligand or at concentrations of either 100 or 200 mM *N*-acetyl-DL-methionine. SYPRO orange was the fluorescent dye. T_m was determined in triplicate in 1°C min^{-1} steps from 20 to 95°C.

6.20, X-ray crystallography of NAAAR G291D F323Y:

6.20.1, Data collection:

Pure NAAAR G291D F323Y (See section 6.10.2) which had been complexed previously with 10 mM *N*-acetyl-DL-methionine, was subjected to crystallization trials using a range of commercially available screens in 96-well plates using 300 nL drops at a range of protein concentrations. The ligand:protein complex was formed by mixing NAAAR with 100 mM ligand (100 mM Tris-HCl (pH 8.0)) to give a final concentration of 10 mM ligand. The best crystals were obtained overnight using CSS A1, A12, B12, and C7 at 20°C, and at a protein concentration of 8 mg mL^{-1}. Larger crystals for diffraction analysis were prepared using the hanging-drop vapour diffusion method in 24-well plate Linbro dishes and using crystallisation drops of 2 µL, with a protein concentration of 8 mg mL^{-1} that had been incubated previously with 10 mM *N*-acetyl-DL-methionine. The best crystals were routinely obtained in crystal drops containing 100 mM Tris-HCl (pH 8.0) 15% PEG 4K, 800 mM sodium

formate. Crystals were tested for diffraction using a Rigaku Micromax-007HF fitted with Osmic multilayer optics and a MARRESEARCH MAR345 imaging plate detector. Those crystals that diffracted to greater than 3 Å resolution were flash-cooled in liquid nitrogen in a cryogenic solution containing the mother liquor and no other cryopotectant, prior to data collection at the synchrotron.

6.20.2, Data collection and processing, structure solution and refinement:

All data collection, processing, structure solutions, and refinement were performed by Dr Gideon Grogan, The University of York. A complete dataset was collected on beamline io-4 at the Diamond Light Source Synchrotron in Oxford on 11th July 2011. Data were processed on the beamline using the expert data processing system Xia2 and the data collection statistics are given in (**Table 19**)[273]. The crystals had the space group *H*32, the same as that described for the WT NAAAR crystals (1SJA)[100]. Coordinates for one monomer of the NAAAR tetramer, derived from PDB file 1SJA were used as a molecular replacement model for structure solution of the G291D F323Y variant using PHASER The solution contained four molecules in the asymmetric unit, the same as for 1SJA[100, 274]. The model was improved using iterative rounds of building in Coot and refinement using REFMAC[275, 276]. In refinement, non-crystallographic symmetry (NCS) restraints were applied for chains B, C and D against chain A. After building the protein and water molecules, clear electron density was observed to persist throughout refinement in the active site of the enzyme, as observed in the *Fo-Fc* map at a level of 4σ. A coordinate file for *N*-acetyl methionine was derived from 1SJA and this ligand was successfully modelled into the density in each subunit and refined using an occupancy level of 1.0. The final ligand-bound structure exhibited R_{cryst} and R_{free} values of 14.8 and 20.3% respectively and was finally validated using PROCHECK[277]. Refinement statistics are presented in (**Table 19**). The Ramachandran plot showed 95.3% of residues to be situated in the most favoured regions, 3.8% in additional allowed and 0.9% in generously allowed regions.

Table 19, Data collection and refinement statistics for NAAAR G291D F323Y variant. Figures in brackets refer to the highest resolution shell.

	NAAAR G291D F323Y complexed with *N*-acetyl methionine
Beamline	Diamond io-3
Wavelength (Å)	0.97630 Å
Space Group	H32
Unit cell (Å)	a = b = 216.71; c = 261.10
Resolution	88.31-2.71 (2.78-2.71)
Unique Reflections	60693
Completeness (%)	100 (100)
Rmerge	0.085 (0.503)
Rp.i.m.	0.041 (0.241)
Multiplicity	10.2 (10.4)
$\langle I/\sigma(I)\rangle$	18.6 (4.8)
Protein atoms	10967
Rcryst/Rfree (%)	14.8/20.3
r.m.s.d. 1-2 bonds (Å)	0.020
r.m.s.d. 1-3 bonds (°)	2.12
Average B factor (Å2)	50

6.21 Protein sequences and alignments:

6.21.1, N-acetyl amino acid racemase (Amycolatopsis Ts-1-60):

GenBank: BAA06400.1

Accession: BAA06400

Protein sequence with G291 and F323 shown in red:

```
1   MKLSGVELRR VQMPLVAPFR TSFGTQSVRE LLLLRAVTPA GEGWGECVTM AGPLYSSEYN

61  DGAEHVLRHY LIPALLAAED ITAAKVTPLL AKFKGHRMAK GALEMAVLDA ELRAHERSFA

121 AELGSVRDSV PCGVSVGIMD TIPQLLDVVG GYLDEGYVRI KLKIEPGWDV EPVRAVRERF

181 GDDVLLQVDA NTAYTLGDAP QLARLDPFGL LLIEQPLEEE DVLGHAELAR RIQTPICLDE

241 SIVSARAAAD AIKLGAVQIV NIKPGRVGGY LEARRVHDVC AAHGIPVWCG GMIETGLGRA

301 ANVALASLPN FTLPGDTSAS DRFYKTDITE PFVLSGGHLP VPTGPGLGVA PIPELLDEVT

361 TAKVWIGS*
```

6.21.2, Sequence alignment of N-acetyl amino acid racemase and L-Ala-D/L-Glu epimerase from Amycolatopsis mediterranei U32.

The five conserved catalytic residues (K163, K263, D189, E214, and D239, numbering is relative to *Am. Ts-1-60* NAAAR) are shown in yellow.

```
NAAAR_          --------MKLTGVELLRVRMPLVAPFRTSFGTQAER--ELLLLRAVTSEGEGWGECGAM 50
AEE_            MTGAPSAGDRADPAEFVVDRVEAWAIALPAVRSLERRTAPRILVKVTAGGVTGWGEA--- 57
                        :   .*::  *:   *   .:. :  .*    :*::..:.   ****.

NAAAR_          AGPLYSAEYIDGVEHVLQAFLVPALLAAGDLTAHKVAPLLAKFK------GHRMAKAALE 104
AEE_            -TPARTCETAESILATIRYHLAPALVGRSAWDLDGITSAFDRAVNCGFTIGAPPAKAAVD 116
                  *  :.*  :.:  .:: .*.***:. .    . ::. : :        *   ****::

NAAAR_          MAVLDAELRAHGLSFGTALGS-TSDSVPCGVSVGIMDSIPQLVDVVGGYLDAGYLRIKLK 163
AEE_            MAVHDLLGRALGLSLGVLWGARRTDRLELGWTVAGR-SAAEAADSVAEGLDHGYRAFKVD 175
                *** *   ** ***:*.  *:  :* :  * :*.   * .: .* *.  ** **  :*:.

NAAAR_          IEPGWDVEPVRAVRER--FGDDILLQVDANTAYTLSDVPQLQR-LDPFGLLLIEQPLEEE 220
AEE_            LGGLGRREDAGIVEAVRAAASDAVLWADGNQSYTVDSALRVAGPLADVGATVFEQPLPAN 235
                :      * .  *.     ..* :* .*.* :**:... ::   *  .*  ::****  :

NAAAR_          DVLGHAELARHLRTPICLDETVVSAASAAAAIRLGACRIVNIKPSRVGGYLEARRVHDVC 280
AEE_            DIAGLKRLRDASPVPIALDEGLPHPSDLATLVRLDAVDVVVAKVQRSGGLTLSRRLCGFA 295
                *: *  .*     .**.*** :  .:. *: :**.*  :*  * .* **   :**: ...

NAAAR_          AAHGIPVWCGGMLETGLGRAANVALASLPGFTLPGDTSASDRFYRTDIT-EPFVLEAGRL 339
AEE_            EDCGTRLMGSGLADSDLGLAASLHLFAAYDIGTPVDLAGRQFVRSPYATGRTVTVVDGAA 355
                   *  :  .*: ::.** **.: * :  .:  * * :. : .  .  * ....:  *

NAAAR_          PVPSGPGLGVTPIPERLAEVTTAKSWLA 367
AEE_            EVPTGPGLGVDVDEQVVRSLAVDLPG-- 381
                 **:******    : : .::.  .
```

6.21.3, Sequence NAAAR from Am. Azurea, Am. Lurida, Am. Ts-1-60, Am. Mediterranei, and G. kaustophilus.

The five conserved catalytic residues (K163, K263, D189, E214, and D239, numbering is relative to *Am. Ts-1-60* NAAAR) are shown in yellow. The location of G291, R299 and F323 are shown in cyan.

```
Azurea_         --MKLSGVELRRVRMPLVAPFRTSFGTQSERELMLVRAVTPAGE-GWGECVTMAAPVYSS  57
Lurida_         --MKLSGVELRRVRMPLVAPFRTSFGTQSERELLLVRAVTPAGE-GWGECVAMEAPLYSS  57
Ts-1-60_        --MKLSGVELRRVQMPLVAPFRTSFGTQSVRELLLLRAVTPAGE-GWGECVTMAGPLYSS  57
Med_            --MKLTGVELLRVRMPLVAPFRTSFGTQAERELLLLRAVTSEGE-GWGECGAMAGPLYSA  57
Geo_            MAINIEYVILRHLQMELKAPFTTSFGTFQTKEFILVEVVDCDGVSGWGESVAFSVPWYSE  60
                  :::  * * :::* * *** *****   :*::*:..*   *  ****. ::  * **

Azurea_         EYNDAAEHVLRNHLIPALLAADDVTAYKVTPLLAKFKGHRMAKGALEMAVLDAELRAHER 117
Lurida_         EYNDAAEHVLRNHLIPALLAAEDVTAHKVTPLLAKFKGHRMAKGALEMAVLDAELRAHDR 117
Ts-1-60_        EYNDGAEHVLRHYLIPALLAAEDITAAKVTPLLAKFKGHRMAKGALEMAVLDAELRAHER 117
Med_            EYIDGVEHVLQAFLVPALLAAGDLTAHKVAPLLAKFKGHRMAKAALEMAVLDAELRAHGL 117
Geo_            ETVKTNWHMLEEFLVPLLFSKPLRHPAELPERFAAIRQNNMAKAALEGAVWDLYAKRLGV 120
                *  .   *:*. .*:* *::     . ::.  :* :: :.***.*** ** *   :

Azurea_         SFAAELGSTRDSVACGVSVGIMDSIPQLLDVVGDYLDEGYVRIKLKIEPGWDIEPVRQVR 177
Lurida_         SFAAELGSTRDSVACGVSVGIMDSIPHLLDVVGGYLDEGYVRIKLKIEPGWDVEPVRQVR 177
Ts-1-60_        SFAAELGSVRDSVPCGVSVGIMDTIPQLLDVVGGYLDEGYVRIKLKIEPGWDVEPVRAVR 177
Med_            SFGTALGSTSDSVPCGVSVGIMDSIPQLVDVVGGYLDAGYLRIKLKIEPGWDVEPVRAVR 177
Geo_            PLCQALGGTKKEIEVGVSIGIQPTVDDLLQVIERYVAQGYRRIKVKIKPGWDVDVIRDVR 180
                .:   **.. ..:  ***:**  :: .*::*:  *:  ** ***:**:****:: :* **

Azurea_         ERFGDDVLLQVDANTAYTLGDAPLLARLDPFDLLLIEQPLEEEDVHGHAELAKRIRTPIC 237
Lurida_         ERFGDDVLLQVDANTAYTLGDAPLLSRLDPFDLLLIEQPLEEEDVLGHAELAKRIRTPIC 237
Ts-1-60_        ERFGDDVLLQVDANTAYTLGDAPQLARLDPFGLLLIEQPLEEEDVLGHAELARRIQTPIC 237
Med_            ERFGDDILLQVDANTAYTLSDVPQLQRLDPFGLLLIEQPLEEEDVLGHAELARHLRTPIC 237
Geo_            RAFP-DVPLMADANSAYTLADAKRLQALDEFGLMMIEQPLAADDLVDHARLQPLLKTPIC 239
                . *  *: * .***:****.*.  *  ** *.*::***** :*: .**.*   ::****

Azurea_         LDESIVSAKAAADAIKLGACQIVNIKPGRVGGYLEARRVHDVCAAHGVAVWCGGMIETGL 297
Lurida_         LDESIVSAKAAADAIKLGACQIVNIKPGRVGGYLEARRVHDVCAAHGIAVWCGGMIETGL 297
Ts-1-60_        LDESIVSARAAADAIKLGAVQIVNIKPGRVGGYLEARRVHDVCAAHGIPVWCGGMIETGL 297
Med_            LDETVVSAASAAAAIRLGACRIVNIKPSRVGGYLEARRVHDVCAAHGIPVWCGGMLETGL 297
Geo_            LDESIRSYDDARKALDLGSCRIINIKIGRVGGLWEAKRIHDLCAERGVPVWCGGMLEAGV 299
                ***:: *   *  *: **: :*:*** .****  **:*:**:** :*:.******:*:*:

Azurea_         GRAANVALASLPGFTLPGDTSASGRFYRTDITEPFVLEA-GHLPVPTGPGLGVTPIPDIL 356
Lurida_         GRAANVALASLPGFTLPGDTSASGRFYRTDITEPFVLDA-GHLPVPTGPGLGVTPIPDLL 356
Ts-1-60_        GRAANVALASLPNFTLPGDTSASDRFYKTDITEPFVLSG-GHLPVPTGPGLGVAPIPELL 356
Med_            GRAANVALASLPGFTLPGDTSASDRFYRTDITEPFVLEA-GRLPVPSGPGLGVTPIPERL 356
Geo_            GRAHNIAITTLENFALPGDTAASSHYWERDIITPEVEVHNGLIRVPNAPGIGYDVDRRQV 359
                *** *:*:::* .*:*****:**.::::. **  * *    * : **..**:*       :

Azurea_         DDFTTEKVWIGS---- 368
Lurida_         DEVTTEKAWIGS---- 368
Ts-1-60_        DEVTTAKVWIG----- 367
Med_            AEVTTAKSWLA----- 367
Geo_            ERYTQFAKLFHRTATA 375
                   *     :
```

7.1, References:

[1] D. Cavalieri, P. E. McGovern, D. L. Hartl, R. Mortimer, M. Polsinelli, *J. Mol. Evol.* **2003**, *57 Suppl 1*, S226.

[2] A. Ghamen, *Tetrahedron* **2007**, *63*, 1721.

[3] M. K. Turner, *Trends Biotechnol.* **1995**, *13*, 173.

[4] A. L. Demain, *J. Ind. Microbiol. Biotechnol.* **2001**, *27*, 352.

[5] S. J. Chiang, *J. Ind. Microbiol. Biotechnol.* **2004**, *31*, 99.

[6] J. B. van Beilen, Z. Li, *Curr. Opin. Biotechnol.* **2002**, *13*, 338.

[7] L. Hilterhaus, A. Liese, *Adv. Biochem. Eng. Biotechnol.* **2007**, *105*, 133.

[8] M. Alcalde, M. Ferrer, P. F. J., A. Ballesteros, *Trends Biotechnol.* **2006**, *24*, 281.

[9] N. Ran, L. Zhao, Z. Chen, J. Tao, *Green Chem.* **2008**, *10*, 361.

[10] C. Delhomme, D. Weuster-Botz, F. E. Kuhn, *Green Chem.* **2008**, *11*, 13.

[11] R. McKenna, D. R. Nielsen, *Metab. Eng.* **2011**, *13*, 544.

[12] T. Willke, K.-D. Vorlop, *Appl. Microbiol. Biotechnol.* **2004**, *66*, 131.

[13] A. J. Straathof, S. Panke, A. Schmid, *Curr. Opin. Biotechnol.* **2002**, *13*, 548.

[14] T. Nagasawa, H. Shimizu, H. Yamada, *Appl. Microbiol. Biotechnol.* **1993**, *40*, 189.

[15] H. Yamada, M. Kobayashi, *Biosci. Biotechnol. Biochem.* **1996**, *60*, 1391.

[16] F. H. Arnold, A. A. Volkov, *Curr. Opin. Chem. Biol.* **1999**, *3*, 54.

[17] N. J. Turner, *Nat. Chem. Biol.* **2009**, *5*, 567.

[18] C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman, G. J. Hughes, *Science* **2010**, *329*, 305.

[19] R. Singh, R. Sharma, N. Tewari, D. S. Rawat, *Chem. Biodivers.* **2006**, *3*, 1279.

[20] L. Martinkova, V. Kren, *Curr. Opin. Chem. Biol.* **2010**, *14*, 130.

[21] J. A. Chaplin, M. D. Levin, B. Morgan, N. Farid, J. Li, Z. Zhu, J. McQuaid, L. W. Nicholson, C. A. Rand, M. J. Burk, *Tetrahedron: Asymmetry* **2004**, *15*, 2793.

[22] Y. Asano, Y. Kato, C. Levy, P. Baker, D. Rice, *Biocatal. Biotransfor.* **2004**, *22*, 131.

[23] N. J. Turner, *Curr. Opin. Chem. Biol.* **2011**, *15*, 234.

[24] M. J. van der Werf, W. J. van den Tweel, J. Kamphuis, S. Hartmans, J. A. de Bont, *Trends Biotechnol.* **1994**, *12*, 95.

[25] B. Weiner, G. J. Poelarends, D. B. Janssen, B. L. Feringa, *Chemistry* **2008**, *14*, 10094.

[26] S. Yamada, K. Nabe, N. Izuo, K. Nakamichi, I. Chibata, *Appl. Environ. Microbiol.* **1981**, *42*, 773.

[27] T. F. Schwede, J. Retey, G. E. Schulz, *Biochemistry* **1999**, *38*, 5355.

[28] N. P. Botting, M. Akhtar, M. A. Cohen, D. Gani, *Biochemistry* **1988**, *27*, 2953.

[29] M. T. Reetz, *Curr. Opin. Chem. Biol.* **2002**, *6*, 145.

[30] A. Zaks, A. M. Klibanov, *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 3192.

[31] J. Y. Houng, M. L. Wu, S. T. Chen, *Chirality* **1996**, *8*, 418.

[32] S. V. Malhotra, H. Zhao, *Chirality* **2005**, *17 Suppl.*, S240.

[33] Y. K. Choi, Y. Kim, K. Han, J. Park, M. J. Kim, *J. Org. Chem.* **2009**, *74*, 9543.

[34] N. J. Turner, *Curr. Opin. Chem. Biol.* **2004**, *8*, 114.

[35] M. Alexeeva, R. Carr, N. J. Turner, *Org. Biomol. Chem.* **2003**, *1*, 4133.

[36] R. Carr, M. Alexeeva, A. Enright, T. S. Eve, M. J. Dawson, N. J. Turner, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 4807.

[37] P. D. Pawelek, J. Cheah, R. Coulombe, P. Macheroux, S. Ghisla, A. Vrielink, *The EMBO journal* **2000**, *19*, 4204.

[38] S. Umhau, L. Pollegioni, G. Molla, K. Diederichs, W. Welte, M. S. Pilone, S. Ghisla, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 12463.

[39] V. I. Tishkov, S. V. Khoronenkova, *Biochemistry (Mosc)* **2005**, *70*, 40.

[40] F. R. Alexandre, D. P. Pantaleone, P. P. Taylor, I. G. Fotheringham, D. J. Ager, N. J. Turner, *Tetrahedron Letters* **2002**, *43*, 707.

[41] R. Fernández-Lafuente, V. Rodriguez, J. M. Guisán, *Enzyme Microb. Technol.* **1998**, *23*, 28.

[42] H. Luo, Q. Li, H. Yu, Z. Shen, *Biotechnol. Lett.* **2004**, *26*, 939.
[43] S. K. Mujawar, *Hindustan. Antibiot. Bull.* **1999**, *41*, 1.
[44] L. Pollegioni, L. Caldinelli, G. Molla, S. Sacchi, M. S. Pilone, *Biotechnol. Prog.* **2004**, *20*, 467.
[45] D. Monti, G. Carrea, S. Riva, E. Baldaro, G. Frare, *Biotechnol. Bioeng.* **2000**, *70*, 239.
[46] S. M. De Wildeman, T. Sonke, H. E. Schoemaker, O. May, *Acc. Chem. Res.* **2007**, *40*, 1260.
[47] F. Paradisi, P. A. Conway, A. R. Maguire, P. C. Engel, *Org. Biomol. Chem.* **2008**, *6*, 3611.
[48] K. Vedha-Peters, M. Gunawardana, J. D. Rozzell, S. J. Novick, *J. Am. Chem. Soc.* **2006**, *128*, 10923.
[49] A. Menzel, H. Werner, J. Altenbuchner, H. Groger, *Eng. Life Sci.* **2004**, *4*, 573.
[50] W. Hummel, M. Kuzu, B. Geueke, *Org. Lett.* **2003**, *5*, 3649.
[51] D. Koszelewski, K. Tauber, K. Faber, W. Kroutil, *Trends Biotechnol.* **2010**, *28*, 324.
[52] J. E. Ayling, H. C. Dunathan, E. E. Snell, *Biochemistry* **1968**, *7*, 4537.
[53] M. D. Truppo, J. D. Rozzell, J. C. Moore, N. J. Turner, *Org. Biomol. Chem.* **2009**, *7*, 395.
[54] K. Yonaha, S. Toyama, H. Kagamiyama, *J. Biol. Chem.* **1983**, *258*, 2260.
[55] M. D. Truppo, N. J. Turner, J. D. Rozzell, *Chem. Comm. (Camb)* **2009**, 2127.
[56] P. Tufvesson, J. Lima-Ramos, J. S. Jensen, N. Al-Haque, W. Neto, J. M. Woodley, *Biotechnol. Bioengin.* **2011**, *108*, 1479.
[57] D. Koszelewski, I. Lavandera, D. Clay, G. M. Guebitz, D. Rozzell, W. Kroutil, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 9337.
[58] T. Eguchi, Y. Kuge, K. Inoue, N. Yoshikawa, K. Mochida, T. Uwajima, *Biosci. Biotechnol. Biochem.* **1992**, *56*, 701.
[59] J. M. Clemente-Jimenez, S. Martinez-Rodriguez, F. Rodriguez-Vico, F. J. Heras-Vazquez, *Recent. Pat. Biotechnol.* **2008**, *2*, 35.
[60] D. C. Lee, G. J. Kim, Y. K. Cha, C. Y. Lee, H. S. Kim, *Biotechnol. Bioeng.* **1997**, *56*, 449.
[61] J. Altenbuchner, M. Siemann-Herzberg, C. Syldatk, *Curr. Opin. Biotechnol.* **2001**, *12*, 559.
[62] C. H. Kao, H. H. Lo, S. K. Hsu, W. H. Hsu, *J. Biotechnol.* **2008**, *134*, 231.
[63] H. H. Lo, C. H. Kao, D. S. Lee, T. K. Yang, W. H. Hsu, *Chirality* **2003**, *15*, 699.
[64] M. Battilotti, U. Barberini, *J .Mol. Catal.* **1988**, *43*, 343.
[65] D. C. Lee, H. S. Kim, *Biotechnol. Bioeng.* **1998**, *60*, 729.
[66] S.-K. Hsu, H.-H. Lo, W.-D. Lin, I.-C. Chen, C.-H. Kao, W.-H. Hsu, *Process Biochem.* **2007**, *42*, 856.
[67] M. Kobayashi, M. Goda, S. Shimizu, *FEBS Lett.* **1998**, *439*, 325.
68] G. Volpato, R. C. Rodrigues, R. Fernandez-Lafuente, *Curr. Med. Chem.* **2010**, *17*, 3855.
[69] J. M. Hamilton-Miller, *Int. J. Antimicrob. Agents* **2008**, *31*, 189.
[70] M. Arroyo, I. de la Mata, C. Acebal, M. P. Castillon, *Appl. Microbiol. Biotechnol.* **2003**, *60*, 507.
[71] J. Rajendhran, P. Gunasekaran, *J. Biosci. Bioeng.* **2004**, *97*, 1.
[72] Y. Kim, W. G. Hol, *Chem. Biol.* **2001**, *8*, 1253.
[73] A. K. Chandel, L. Venkateswar Rao, M. Lakshmi Narasu, O. V. Singh, *Enzyme Micro. Tech.* **2008**, *42*, 199.
[74] R. V. Ulijn, L. De Martin, P. J. Halling, B. D. Moore, A. E. Janssen, *J. Biotechnol.* **2002**, *99*, 215.
[75] M. I. Youshko, L. M. van Langen, E. de Vroom, F. van Rantwijk, R. A. Sheldon, V. K. Svedas, *Biotechnol. Bioeng.* **2002**, *78*, 589.
[76] Y. Asano, S. Yamaguchi, *J Am Chem Soc* **2005**, *127*, 7696.
[77] S. Yamaguchi, H. Komeda, Y. Asano, *Appl. Environ. Microbiol.* **2007**, *73*, 5370.
[78] K. Plhackova, V. Vojtisek, J. Plachy, *Folia. Microbiol. (Praha)* **1982**, *27*, 382.

[79] T. Sonke, S. Ernste, R. F. Tandler, B. Kaptein, W. P. Peeters, F. B. van Assema, M. G. Wubbolts, H. E. Schoemaker, *Appl. Environ. Microbiol.* **2005**, *71*, 7961.
[80] C. Carboni, H. G. T. Kierkels, L. Gardossi, K. Tamiola, J. D. B., P. J. L. M. Quaedfliega, *Tetrahedron: Asymmetry* **2006**, *17*, 245.
[81] D. H. Baek, S. J. Kwon, S. P. Hong, M. S. Kwak, M. H. Lee, J. J. Song, S. G. Lee, K. H. Yoon, M. H. Sung, *Appl. Environ. Microbiol.* **2003**, *69*, 980.
[82] A. S. Bommarius, K. Drauz, H. Klenk, C. Wandrey, *Ann. N. Y. Acad. Sci.* **1992**, *672*, 126.
[83] L. A. Mounter, L. T. Dien, F. E. Bell, *J. Biol. Chem.* **1958**, *233*, 903.
[84] F. E. Bell, L. A. Mounter, *J. Biol. Chem.* **1958**, *233*, 900.
[85] H. K. Chenault, J. Dahmer, G. M. Whitesides, *J. Am. Chem. Soc.* **1989**, *111*, 6354.
[86] H. S. Toogood, E. J. Hollingsworth, R. C. Brown, I. N. Taylor, S. J. C. Taylor, R. McCague, J. A. Littlechild, *Extremophiles* **2002**, *6*, 111.
[87] K. Sakai, T. Obata, K. Ideta, M. Moriguchi, *J. Ferment. Bioeng.* **1991**, *71*, 79.
[88] S. V. Story, A. M. Grunden, M. W. Adams, *J. Bacteriol.* **2001**, *183*, 4259.
[89] O. May, S. Verseck, A. Bommarius, K. Drauz, *Org. Process. Res. Dev.* **2002**, *6*, 452.
[90] T. C. Stadtman, P. Elliott, *J. Biol. Chem.* **1957**, *228*, 983.
[91] S. Tokuyama, K. Hatano, *Appl. Microbiol. Biotechnol.* **1996**, *44*, 774.
[92] S. K. Hsu, H. H. Lo, C. H. Kao, D. S. Lee, W. H. Hsu, *Biotechnol. Prog.* **2006**, *22*, 1578.
[93] S. Tokuyama, K. Hatano, T. Takahashi, *Biosci. Biotech. Bioch.* **1994**, *58*, 24.
[94] S. Tokuyama, K. Hatano, *Appl. Microbiol. Biotechnol.* **1995**, *42*, 884.
[95] S. Tokuyama, K. Hatano, *Appl. Microbiol. Biotechnol.* **1995**, *42*, 853.
[96] S. Shiun-Cheng, L. Chia-Yin, *Enzyme Micro Tech* **2002**, *30*, 647.
[97] S. Verseck, A. Bommarius, M. R. Kula, *Appl. Microbiol. Biotechnol.* **2001**, *55*, 354.
[98] W. C. Wang, W. C. Chiu, S. K. Hsu, C. L. Wu, C. Y. Chen, J. S. Liu, W. H. Hsu, *J. Mol. Biol.* **2004**, *342*, 155.
[99] M. Hayashida, S. H. Kim, K. Takeda, T. Hisano, K. Miki, *Proteins* **2008**, *71*, 519.
[100] J. B. Thoden, E. A. Taylor Ringia, J. B. Garrett, J. A. Gerlt, H. M. Holden, I. Rayment, *Biochemistry* **2004**, *43*, 5716.
[101] E. A. Taylor Ringia, J. B. Garrett, J. B. Thoden, H. M. Holden, I. Rayment, J. A. Gerlt, *Biochemistry* **2004**, *43*, 224.
[102] P. C. Babbitt, M. S. Hasson, J. E. Wedekind, D. R. Palmer, W. C. Barrett, G. H. Reed, I. Rayment, D. Ringe, G. L. Kenyon, J. A. Gerlt, *Biochemistry* **1996**, *35*, 16489.
[103] D. R. Palmer, J. B. Garrett, V. Sharma, R. Meganathan, P. C. Babbitt, J. A. Gerlt, *Biochemistry* **1999**, *38*, 4252.
[104] V. A. Klenchin, E. A. Taylor Ringia, J. A. Gerlt, I. Rayment, *Biochemistry* **2003**, *42*, 14427.
[105] M. E. Glasner, N. Fayazmanesh, R. A. Chiang, A. Sakai, M. P. Jacobson, J. A. Gerlt, P. C. Babbitt, *J. Mol. Biol.* **2006**, *360*, 228.
[106] A. Sakai, D. F. Xiang, C. Xu, L. Song, W. S. Yew, F. M. Raushel, J. A. Gerlt, *Biochemistry* **2006**, *45*, 4455.
[107] S. D. Brown, J. A. Gerlt, J. L. Seffernick, P. C. Babbitt, *Genome Biol.* **2006**, *7*, R8.
[108] J. A. Gerlt, P. C. Babbitt, *Annu. Rev. Biochem.* **2001**, *70*, 209.
[109] J. A. Gerlt, P. C. Babbitt, I. Rayment, *Arch. Biochem. Biophys.* **2005**, *433*, 59.
[110] P. C. Babbitt, G. T. Mrachko, M. S. Hasson, G. W. Huisman, R. Kolter, D. Ringe, G. A. Petsko, G. L. Kenyon, J. A. Gerlt, *Science* **1995**, *267*, 1159.
[111] J. Gerlt, *Biochemistry* **2009**.
[112] T. G. Spring, F. Wold, *J. Biol. Chem.* **1971**, *246*, 6797.
[113] C. W. Levy, P. A. Buckley, S. Sedelnikova, Y. Kato, Y. Asano, D. W. Rice, P. J. Baker, *Structure* **2002**, *10*, 105.
[114] J. A. Gerlt, F. M. Raushel, *Curr. Opin. Chem. Biol.* **2003**, *7*, 252.
[115] J. F. Rakus, A. A. Fedorov, E. V. Fedorov, M. E. Glasner, J. E. Vick, P. C. Babbitt, S. C. Almo, J. A. Gerlt, *Biochemistry* **2007**, *46*, 12896.

[116] D. M. Schmidt, E. C. Mundorff, M. Dojka, E. Bermudez, J. E. Ness, S. Govindarajan, P. C. Babbitt, J. Minshull, J. A. Gerlt, *Biochemistry* **2003**, *42*, 8387.
[117] J. E. Vick, D. M. Schmidt, J. A. Gerlt, *Biochemistry* **2005**, *44*, 11722.
[118] J. E. Vick, J. A. Gerlt, *Biochemistry* **2007**, *46*, 14589.
[119] D. Reardon, G. K. Farber, *Faseb. J.* **1995**, *9*, 497.
[120] A. R. Raine, N. S. Scrutton, F. S. Mathews, *Protein Sci.* **1994**, *3*, 1889.
[121] K. M. Fox, P. A. Karplus, *Structure* **1994**, *2*, 1089.
[122] M. Henn-Sax, B. Hocker, M. Wilmanns, R. Sterner, *Biol. Chem.* **2001**, *382*, 1315.
[123] D. M. Schmidt, B. K. Hubbard, J. A. Gerlt, *Biochemistry* **2001**, *40*, 15707.
[124] J. Pozo-Dengra, A. I. Martínez-Gómeza, S. Martínez-Rodrígueza, J. M. Clemente-Jiméneza, F. Rodríguez-Vicoa, L. H.-V. F.J., *Process Biochem.* **2009**, *44*, 835.
[125] J. Pozo-Dengra, S. Martinez-Rodriguez, L. M. Contreras, J. Prieto, M. Andujar-Sanchez, J. M. Clemente-Jimenez, F. J. Las Heras-Vazquez, F. Rodriguez-Vico, J. L. Neira, *Biopolymers* **2009**, *91*, 757.
[126] T. J. Lupoli, H. Tsukamoto, E. H. Doud, T. S. Wang, S. Walker, D. Kahne, *J. Am. Chem. Soc.* ***2011***, *133*, 10748.
[127] S. K. Hsu, H. H. Loc, W. D. Lina, I. C. Chena, C. H. Kaoe, W. H. Hsu, *Process. Biochem.* **2007**, *42*, 856.
[128] T. B. Thompson, J. B. Garrett, E. A. Taylor, R. Meganathan, J. A. Gerlt, I. Rayment, *Biochemistry* **2000**, *39*, 10662.
[129] J. A. Gerlt, P. G. Gassman, *Biochemistry* **1993**, *32*, 11943.
[130] F. Cava, H. Lam, M. A. de Pedro, M. K. Waldor, *Cell. Mol. Life Sci.* **2011**, *68*, 817.
[131] V. S. Lamzin, Z. Dauter, K. S. Wilson, *Curr. Opin. Struc. Bio.* **1995**, *5*, 830.
[132] T. Yoshimura, M. Goto, *FEBS J.* **2008**, *275*, 3527.
[133] J. Lee, D. G. Lee, *Exp. Mol. Med.* **2008**, *40*, 370.
[134] M. L. Mangoni, N. Papo, J. M. Saugar, D. Barra, Y. Shai, M. Simmaco, L. Rivas, *Biochemistry* **2006**, *45*, 4266.
[135] M. R. Levengood, C. C. Kerwood, C. Chatterjee, W. A. van der Donk, *Chembiochem* **2009**, *10*, 911.
[136] H. E. Umbarger, *Annu. Rev. Biochem.* **1978**, *47*, 532.
[137] A. Watanabe, T. Yoshimura, B. Mikami, H. Hayashi, H. Kagamiyama, N. Esaki, *J. Biol. Chem.* **2002**, *277*, 19166.
[138] S. Y. Choi, N. Esaki, T. Yoshimura, K. Soda, *J. Biochem.* **1992**, *112*, 139.
[139] S. Glavas, M. E. Tanner, *Biochemistry* **1999**, *38*, 4106.
[140] T. Yamauchi, S. Y. Choi, H. Okada, M. Yohda, H. Kumagai, N. Esaki, K. Soda, *J. Biol. Chem.* **1992**, *267*, 18361.
[141] K. Y. Hwang, C. S. Cho, S. S. Kim, H. C. Sung, Y. G. Yu, Y. Cho, *Nat. Struct. Biol.* **1999**, *6*, 422.
[142] M. A. Spies, J. G. Reese, D. Dodd, K. L. Pankow, S. R. Blanke, J. Baudry, *J. Am. Chem. Soc.* **2009**, *131*, 5274.
[143] A. Buschiazzo, M. Goytia, F. Schaeffer, W. Degrave, W. Shepard, C. Gregoire, N. Chamond, A. Cosson, A. Berneman, N. Coatnoan, P. M. Alzari, P. Minoprio, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 1705.
[144] M. Kanda, K. Hori, T. Kurotsu, S. Miura, Y. Saito, *J. Biochem.* **1989**, *105*, 653.
[145] C. L. Chen, L. K. Chang, Y. S. Chang, S. T. Liu, J. S. Tschen, *Mol. Gen. Genet.* **1995**, *248*, 121.
[146] S. G. Laland, O. Froyshov, C. Gilhuus-Moe, T. L. Zimmer, *Nat. New. Biol.* **1972**, *239*, 43.
[147] B. Marrs, S. Delagrave, D. Murphy, *Curr. Opin. Microbiol* .**1999**, *2*, 241.
[148] J. D. Rozzell, *Bioorg. Med. Chem.* **1999**, *7*, 2253.
[149] D. Georlette, V. Blaise, T. Collins, S. D'Amico, E. Gratia, A. Hoyoux, J. C. Marx, G. Sonan, G. Feller, C. Gerday, *FEMS Microbiol. Rev.* **2006**, *28*, 25.
[150] J. L. Jestin, S. Vichier-Guerre, *Res. Microbiol.* **2005**, *156*, 961.

[151] K. A. Powell, S. W. Ramer, S. B. Del Cardayre, W. P. Stemmer, M. B. Tobin, P. F. Longchamp, G. W. Huisman, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 3948.
[152] O. Khersonsky, C. Roodveldt, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2006**, *10*, 498.
[153] H. Kunz, *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 4439.
[154] M. Adamczak, S. H. Krishna, *Food Technol. Biotechnol.* **2004**, *42*, 251.
[155] A. Glieder, E. T. Farinas, F. H. Arnold, *Nat. Biotechnol.* **2002**, *20*, 1135.
[156] R. Fasan, Y. T. Meharenna, C. D. Snow, T. L. Poulos, F. H. Arnold, *J. Mol. Biol.* **2008**, *383*, 1069.
[157] C. Neylon, *Nucleic Acids Res.* **2004**, *32*, 1448.
[158] M. T. Reetz, D. Kahakeaw, R. Lohmer, *Chembiochem* **2008**, *9*, 1797.
[159] J. M. Smith, *Nature* **1970**, *225*, 563.
[160] M. T. Reetz, L. W. Wang, M. Bocola, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 1236.
[161] T. S. Rasila, M. I. Pajunen, H. Savilahti, *Anal. Biochem.* **2009**, *388*, 71.
[162] T. W. Wang, H. Zhu, X. Y. Ma, T. Zhang, Y. S. Ma, D. Z. Wei, *Mol. Biotechnol.* **2006**, *34*, 55.
[163] A. Greener, M. Callahan, B. Jerpseth, *Methods Mol. Biol.* **1996**, *57*, 375.
[164] R. D. Sleator, C. G. Gahan, C. Hill, *Appl. Environ. Microbiol.* **2001**, *67*, 4560.
[165] P. L. Foster, *Methods Enzymol.* **1991**, *204*, 114.
[166] E. O. McCullum, B. A. Williams, J. Zhang, J. C. Chaput, *Methods Mol. Biol.* **2010**, *634*, 103.
[167] J. H. Spee, W. M. de Vos, O. P. Kuipers, *Nucleic Acids. Res.* **1993**, *21*, 777.
[168] M. Zaccolo, D. M. Williams, D. M. Brown, E. Gherardi, *J. Mol. Biol.* **1996**, *255*, 589.
[169] T. Vanhercke, C. Ampe, L. Tirry, P. Denolf, *Anal. Biochem.* **2005**, *339*, 9.
[170] T. S. Wong, D. Roccatano, M. Zacharias, U. Schwaneberg, *J. Mol. Biol.* **2006**, *355*, 858.
[171] H. Murakami, T. Hohsaka, M. Sisido, *Nat. Biotechnol.* **2002**, *20*, 76.
[172] F. H. Crick, L. Barnett, S. Brenner, R. J. Watts-Tobin, *Nature* **1961**, *192*, 1227.
[173] D. H. Ardell, G. Sella, *J. Mol. Evol.* **2001**, *53*, 269.
[174] T. S. Wong, D. Roccatano, U. Schwaneberg, *Biocatal. Biotransform.* **2007**, *25*, 229.
[175] W. P. Stemmer, *Nature* **1994**, *370*, 389.
[176] W. P. Stemmer, *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 10747.
[177] W. M. Coco, W. E. Levinson, M. J. Crist, H. J. Hektor, A. Darzins, P. T. Pienkos, C. H. Squires, D. J. Monticello, *Nat. Biotechnol.* **2001**, *19*, 354.
[178] G. S. Roeder, *Genes Dev.* **1997**, *11*, 2600.
[179] Y. An, J. Ji, A. Lv, R. Huang, Z. Xiu, *Mol. Biol. (Mosk)* **2006**, *40*, 546.
[180] H. Zhao, F. H. Arnold, *Nucleic Acids Res.* **1997**, *25*, 1307.
[181] A. Crameri, E. A. Whitehorn, E. Tate, W. P. Stemmer, *Nat. Biotechnol.* **1996**, *14*, 315.
[182] L. Zheng, U. Baumann, J. L. Reymond, *Nucleic Acids Res.* **2004**, *32*, e115.
[183] D. L. Steffens, J. G. W. Williams, *J. Biomol. Tech.* **2007**, *18*, 147.
[184] M. T. Reetz, D. Kahakeaw, J. Sanchis, *Mol. Biosyst.* **2009**, *5*, 115.
[185] J. Sanchis, L. Fernandez, J. D. Carballeira, J. Drone, Y. Gumulya, H. Hobenreich, D. Kahakeaw, S. Kille, R. Lohmer, J. J. Peyralans, J. Podtetenieff, S. Prasad, P. Soni, A. Taglieber, S. Wu, F. E. Zilly, M. T. Reetz, *Appl. Microbiol. Biotechnol.* **2008**, *81*, 387.
[186] K. Miyazaki, *Methods Enzymol.* **2011**, *498*, 399.
[187] D. Wei, M. Li, X. Zhang, L. Xing, *Anal. Biochem.* **2004**, *331*, 401.
[188] M. T. Reetz, S. Prasad, J. D. Carballeira, Y. Gumulya, M. Bocola, *J. Am. Chem. Soc.* **2010**, *132*, 9144.
[189] M. T. Reetz, S. Wu, *Chem. Commun. (Camb)* **2008**, 5499.
[190] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891.
[191] M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, A. Vogel, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 4192.
[192] Y. P. Lai, J. Huang, L. F. Wang, J. Li, Z. R. Wu, *Biotechnol. Bioeng.* **2004**, *86*, 622.

[193] J. Nelms, R. M. Edwards, J. Warwick, I. Fotheringham, *Appl. Environ. Microbiol.* **1992**, *58*, 2592.
[194] U. Mohan, U. C. Banerjee, *Chembiochem* **2008**, *9*, 2238.
[195] J. L. Guenet, *Genetica* **2004**, *122*, 9.
[196] Y. An, W. Wu, A. Lv, *Biochimie* **2010**, *92*, 1081.
[197] W. C. Tseng, J. W. Lin, T. Y. Wei, T. Y. Fang, *Anal. Biochem.* **2008**, *375*, 376.
[198] Y. L. Boersma, M. J. Droge, A. M. van der Sloot, T. Pijning, R. H. Cool, B. W. Dijkstra, W. J. Quax, *Chembiochem* **2008**, *9*, 1110.
[199] M. T. Reetz, H. Hobenreich, P. Soni, L. Fernandez, *Chem. Commun. (Camb)* **2008**, 5502.
[200] M. T. Reetz, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 138.
[201] T. Yano, S. Oue, H. Kagamiyama, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 5511.
[202] S. Delagrave, D. J. Murphy, J. L. Pruss, A. M. Maffia, 3rd, B. L. Marrs, E. J. Bylina, W. J. Coleman, C. L. Grek, M. R. Dilworth, M. M. Yang, D. C. Youvan, *Protein Eng.* **2001**, *14*, 261.
[203] M. Gabler, M. Hensel, L. Fischer, *Enzyme Microb. Technol.* **2000**, *27*, 605.
[204] S. Bershtein, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2008**, *12*, 151.
[205] C. Heddle, S. L. Mazaleyrat, *Protein Eng. Des. Sel.* **2007**, *20*, 327.
[206] Y. Ikenaka, H. Nanba, K. Yajima, Y. Yamada, M. Takano, S. Takahashi, *Biosci. Biotechnol. Biochem.* **1998**, *62*, 1672.
[207] H. Kaltwasser, H. G. Schlegel, *Anal. Biochem.* **1966**, *16*, 132.
[208] S. Fong, T. D. Machajewski, C. C. Mak, C. Wong, *Chem. Biol.* **2000**, *7*, 873.
[209] K. Kiianitsa, J. A. Solinger, W. D. Heyer, *Anal. Biochem.* **2003**, *321*, 266.
[210] S. van den Berg, P. A. Lofdahl, T. Hard, H. Berglund, *J. Biotechnol.* **2006**, *121*, 291.
[211] S. Becker, H. U. Schmoldt, T. M. Adams, S. Wilhelm, H. Kolmar, *Curr. Opin. Biotechnol.* **2004**, *15*, 323.
[212] S. Becker, H. Hobenreich, A. Vogel, J. Knorr, S. Wilhelm, F. Rosenau, K. E. Jaeger, M. T. Reetz, H. Kolmar, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 5085.
[213] A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Rothlisberger, D. Baker, *Protein Sci.* **2006**, *15*, 2785.
[214] M. Hohne, S. Schatzle, H. Jochens, K. Robins, U. T. Bornscheuer, *Nat. Chem. Biol.* **2010**, *6*, 807.
[215] K. A. Dill, S. B. Ozkan, T. R. Weikl, J. D. Chodera, V. A. Voelz, *Curr. Opin. Struct. Biol.* **2007**, *17*, 342.
[216] K. A. Dill, S. B. Ozkan, M. S. Shell, T. R. Weikl, *Annu. Rev. Biophys.* **2008**, *37*, 289.
[217] M. A. Dwyer, L. L. Looger, H. W. Hellinga, *Science* **2004**, *304*, 1967.
[218] M. A. Dwyer, L. L. Looger, H. W. Hellinga, *Science* **2008**, *319*, 569.
[219] W. J. Albery, J. R. Knowles, *Biochemistry* **1976**, *15*, 5627.
[220] J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, D. Baker, *Science* **2010**, *329*, 309.
[221] D. Rothlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, *Nature* **2008**, *453*, 190.
[222] K. Auclair, A. Sutherland, J. Kennedy, D. J. Witter, J. P. Van den Heever, C. R. Hutchinson, J. C. Vederas, *J. Am. Chem. Soc.* **2000**, *122*, 11519.
[223] C. R. Guimaraes, M. Udier-Blagovic, W. L. Jorgensen, *J. Am. Chem. Soc.* **2005**, *127*, 3577.
[224] D. Hilvert, K. W. Hill, K. D. Nared, M. T. M. Auditor, *J. Am. Chem. Soc.* **1989**, *111*, 9261.
[225] A. A. Desai, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 1974.
[226] A. Iwasaki, Y. Yamada, N. Kizaki, Y. Ikenaka, J. Hasegawa, *Appl. Microbiol. Biotechnol.* **2006**, *69*, 499.

[227] B. K. Cho, H. Y. Park, J. H. Seo, J. Kim, T. J. Kang, B. S. Lee, B. G. Kim, *Biotechnol. Bioeng.* **2008**, *99*, 275.
[228] J.-S. Shin, B.-G. Kim, *J. Org. Chem.* **2002**, *67*, 2848.
[229] S. Royer, *PhD, The University of Edinburgh, UK* **(2005)**.
[230] M. M. Kaplan, M. Flavin, *Biochim. Biophys. Acta.* **1965**, *104*, 390.
[231] M. M. Kaplan, M. Flavin, *J. Biol. Chem.* **1966**, *241*, 5781.
[232] M. M. Kaplan, M. Flavin, *J. Biol. Chem.* **1966**, *241*, 4463.
[233] J. Wild, T. Klopotowski, *Mol. Gen. Genet.* **1981**, *181*, 373.
[234] J. Wild, M. Filutowicz, T. Klopotowski, *Arch. Microbiol.* **1978**, *118*, 71.
[235] K. A. Datsenko, B. L. Wanner, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 6640.
[236] O. Kuchner, F. H. Arnold, *Trends Biotechnol.* **1997**, *15*, 523.
[237] M. J. Stark, *Gene* **1987**, *51*, 255.
[238] P. R. Jensen, H. V. Westerhoff, O. Michelsen, *Eur. J. Biochem.* **1993**, *211*, 181.
[239] L. Robert, G. Paul, Y. Chen, F. Taddei, D. Baigl, A. B. Lindner, *Mol. Syst. Biol.* **2010**, *6*, 357.
[240] R. S. Donovan, C. W. Robinson, B. R. Glick, *J. Ind. Microbiol.* **1996**, *16*, 145.
[241] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, H. Mori, *Mol. Syst. Biol.* **2006**, *2*, 2006 0008.
[242] R. B. Roberts, P. H. Abelson, D. B. Crowie, E. T. Bolton, R. J. Britten, *Carnegie Inst. Washington Publ.* **1955**, *article 607*.
[243] J. Lederberg, *Methods Med. Res.* **1950**, *3*, 5.
[244] M. Mandel, A. Higa, *J. Mol. Biol.* **1970**, *53*, 159.
[245] W. J. Dower, J. F. Miller, C. W. Ragsdale, *Nucleic Acids Res.* **1988**, *16*, 6127.
[246] X. Shi, T. Karkut, M. Alting-Mees, M. Chamankhah, S. M. Hemmingsen, D. D. Hegedus, *Anal. Biochem.* **2003**, *320*, 152.
[247] S. B. Leslie, E. Israeli, B. Lighthart, J. H. Crowe, L. M. Crowew, *Appl. Enviorn. Mirco.* **1995**, *61*, 3592.
[248] B. T. Storey, E. E. Noiles, K. A. Thompson, *Cryobiology* **1998**, *37*, 46.
[249] S. N. Ho, H. D. Hunt, R. M. Horton, J. K. Pullen, P. L. R., *Gene* **1989**, *77*, 51.
[250] Y. Nakamura, T. Gojobori, T. Ikemura, *Nucleic Acids Res.* **2000**, *28*, 292.
[251] G. von Heijne, *J. Mol. Biol.* **1991**, *218*, 499.
[252] J. Pozo-Dengra, A. I. Martınez-Gomez, S. Martınez-Rodrıguez, J. M. Clemente-Jimenez, F. Rodrıguez-Vico, F. J. Las Heras-Vazquez, *Process Biochem.* **2009**, *44*, 835.
[253] Y. Okamoto, T. Ikai, *Chem. Soc. Rev.* **2008**, *37*, 2593.
[254] D. W. Aswad, *Anal. Biochem.* **1984**, *137*, 405.
[255] D. W. Armstrong, Y. Youbang Liu, K. H. Ekborgott, *Chirality* **1995**, *7*, 474.
[256] Y. P. Yu, S. H. Wu, *Chirality* **2001**, *13*, 231.
[257] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, *Protein Sci.* **1995**, *4*, 2411.
[258] A. Narmandakh, S. L. Bearne, *Protein Expr. Purif.* ***2010***, *69*, 39.
[259] J. Kiraga, P. Mackiewicz, D. Mackiewicz, M. Kowalczuk, P. Biecek, N. Polak, K. Smolarczyk, M. R. Dudek, S. Cebra, *BMC Genomics* **2007**, *8*.
[260] S. C. Su, C. Y. Lee, *Enzyme Microb. Tech.* **2002**, *30*, 647.
[261] J. L. Bada, *Naturwissenschaften* **1975**, *62*, 71.
[262] L. Canoira, M. J. Garcia-martinez, J. F. Llamas, O. J. E., T. D. Torres, *Int. J. Chem. Kinet.* **2003**, *35*, 576.
[263] R. S. Ward, *Tetrahedron: Asymmetry* **1995**, *6*, 1475.
[264] S. Y. Lam, R. C. Yeung, T. H. Yu, K. H. Sze, K. B. Wong, *PLoS Biol.* **2011**, *9*, e1001027.
[265] K. M. Polizzi, A. S. Bommarius, J. M. Broering, J. F. Chaparro-Riggers, *Curr. Opin. Chem. Biol.* **2007**, *11*, 220.
[266] U. B. Ericsson, B. M. Hallberg, G. T. Detitta, N. Dekker, P. Nordlund, *Anal. Biochem.* **2006**, *357*, 289.
[267] S. D'Amico, J. C. Marx, C. Gerday, G. Feller, *J. Biol. Chem.* **2003**, *278*, 7891.
[268] P. A. Fields, G. N. Somero, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 11476.

[269] B. K. Shoichet, W. A. Baase, R. Kuroki, M. B. W., *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 452.
[270] S. Heimbecher, Y. C. Lee, S. E. Tabibi, Y. S. H., *J. Chromatogr. B.* **1997**, *691*, 173.
[271] E. F. Birse, S. A. Eaton, D. E. Jane, P. L. St J. Jones, R. H. P. Porter, P. C. K. Pook, D. C. Sunter, P. M. Udvarhely, B. Wharton, P. J. Roberts, T. F. Salt, J. C. Watkins, *Neuroscience* **1993**, *52*, 481.
[272] R. C. Lloyd, M. C. Lloyd, M. E. B. Smith, K. E. Holt, J. P. Swift, P. A. Keene, S. J. C. Taylor, R. McCague, *Tetrahedron* **2004**, *60*, 717.
[273] J. Winter, *J. Appl. Cryst* **2010**, *43*, 186.
[274] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658.
[275] P. Emsley, K. Cowtan, *Acta. Crystallogr. D. Biol. Crystallogr.* **2004**, *60*, 2126.
[276] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta. Crystallogr. D. Biol. Crystallogr.* **1997**, *53*, 240.
[277] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Cryst.* **1993**, *26*, 283.